A Study Of The Novel Gene Clwd And Its Role In The Pathology Of The Mouse Mutant Wasted.

Jean O'Donoghue

Thesis submitted for the degree of
Doctor of Philosophy
The University of Edinburgh
2004

Table of Contents

List of Figures

Declaration

I declare that this thesis has been composed by myself, and that all the work is my own unless otherwise clearly stated.

Jean O'Donoghue

Acknowledgements

First of all I'd like to thank Cathy for her help, encouragement and most of all constant enthusiasm for a mad little project studying a mad little gene. I feel very lucky to have had a supervisor who not only allowed me go off on a tangent, but who was interested in where I ended up. Next come all past and present members of the lab whose advice and friendship have been invaluable Katie, Dawn, Helen, Julia, Permphan, Vicky and You Ying – thanks for everything! And here I must wish You Ying the best of luck with taking the Clwd story further – I hope you enjoy it as much as I have.
My time in the lab would not have been half so much fun without everyone in Medgen being such fun-loving and above all generous and helpful people. Everyone is so genuine and willing to help, I'm not sure I will ever find a lab quite like it again. Thanks in particular to Heather, Susan, Helen, Ann and Alison for answering too many stupid questions to remember. Everyone at the MRC transgenic unit have been wonderful as well – particularly Tam – for looking after my mice and me! For advice and supplies I also have to thank some fellow Wellcome students, Helen Munn and Katie Brooks and indeed everyone involved in the four-year Wellcome programme –management committee and students alike. I feel very lucky to have been a part of a great programme. I am also indebted to Colin Watts who with Michele and Henrik got me started on dendritic cells and provided their time and expertise so willingly.
Aside from work - I am indebted to all the students in the section for making my time here so enjoyable, in particular to Simon (for movie trailers and mine sweeper), and Sarah (for keeping me sane with late-night chats and counselling sessions). HF must be thanked too for Munro-breaks from the lab but most of all for some wonderful friends. Thanks to my long-distance support team Mum, Dad, Amy and Áine. Between us I think we've kept BT and Eircom in business for the last 4 years and I'm very glad!

Finally, and most importantly, I thank Harry for the love, support and patience without which I would be lost.

Abstract

Wasted (wst) is a recessive spontaneous mutation in the mouse resulting in a complex phenotype incorporating neuromuscular wasting and atrophy of the organs of the immune system. Lymphoid derived cells also display an aberrant response to ionising radiation. The genetic lesion in wasted mice is a 15.8Kb deletion on mouse chromosome 2. This deletion removes the promoter and first non-coding exon of eukaryotic elongation factor 1a2 or eEF1A2. This is a tissue specific isoform of eEF1A1 found in the terminally differentiated cells of the brain, heart and muscle, thus providing a feasible hypothesis for the neuromuscular phenotype. However, how an essential eEF1A2 knockout causes an immunological phenotype remains unclear.

Here I will describe the study of a novel gene in close proximity to the wasted deletion. This new gene Clwd (*Cl*ose to the *W*asted *D*eletion) is a small 1.2kb gene which appears (from EST data) to be ubiquitously expressed. It has several alternative splice forms and a short conserved upstream Open Reading Frame or uORF. The protein it encodes is 115 amino acids long with no striking homology to other proteins in the database. These features have been confirmed using RT-PCR and by generating and using an anti-peptide antibody to Clwd. Immunocytochemistry shows Clwd to be associated with the nuclear membrane. Upon the application of RNA interference (RNAi) to silence the Clwd gene transiently in mammalian cell culture, cells lacking Clwd display dysregulation of the cell cycle with cells appearing to have disrupted nuclei and slowed growth.

As a result of this study I believe it unlikely that Clwd is responsible for the immunological phenotype of wasted mice. There is no difference in Clwd mRNA or protein expression between wasted and wild type controls. A transgenic project has been carried out in the lab whereby mice were injected with a BAC containing the whole genomic region surrounding the deletion, and a BAC containing the whole genomic region, but with eEF1A2 knocked out. Analysis of these animals suggests the pathology

of wasted is due solely to the lack of eEF1A2. I will describe some preliminary studies to assess the importance of eEF1A2 expression in the dendritic cells of the immune system.

Abbreviations/Symbols

4NQO	4-Nitroquinoline 1-oxide
AE	Anchoring Enzyme
AT	Ataxia Telangiectasia
ATM	Ataxia Telangiectasia Mutated
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
bp	base pair(s)
Clwd	Close to wasted deletion
CNS	Central Nervous System
Crp	C-reactive protein, petaxin related
DAB	3,3'-Diaminobenzidine
DABCO	1,4-Diazobicyclo[2,2,2]-octane
DEPC	Diethylpyrocarbonate
DIG	digoxigenin
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
dsRNA	double stranded RNA
DTH	Delayed Type Hypersensitivity
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediamine tetra-acetic acid
Eef1a1	Eukaryotic translation Elongation Factor 1A1 (mouse gene)
EEF1A1	Eukaryotic translation Elongation Factor 1A1 (human gene)
eEF1A1	Eukaryotic translation Elongation Factor 1A1 (mouse protein)
Eef1a2	Eukaryotic translation Elongation Factor 1A1 (mouse gene)
eEF1A2	Eukaryotic translation Elongation Factor 1A1 (mouse protein)

EEF1A2	Eukaryotic translation Elongation Factor 1A2 (human gene)
EIf2A	Eukaryotic translation initiation factor 2A
ELISA	Enzyme Linked ImmunoSorbent Assay
ES cell	Embryonic Stem cell
EST	Expressed Sequence Tag
EtOH	Ethanol
Fn1	Fibronectin 1
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase
gDNA	genomic DNA
GDP	Guanine diphosphate
GPI	Glycosylated phosphatidylinositol
GTP	Guanine triphosphate
HGMP	Human Genome Mapping Project
HSV1-TK	Herpes Simplex Virus type 1 Thymidine Kinase
IAP	Intracisternal A-particle
Il1	Interleukin 1
Il6	Interleukin 6
IMAGE	Integrated Molecular Analysis of Genomes and their Expression
kb	kilobase(s)
kD	kiloDalton(s)
KLH	Keyhole Limpet Haemocyanin
LDAS	Length Difference Alternative Splicing
M	Molar
M_4	Muscarinic acetylcholine receptor 4
MeOH	Methanol
MITF	Microphthalmia-associated transcription factor:
ml	millilitrre
MOPS	3-[N-Morpholino]propanesulphonic acid
mRNA	messenger RNA
mTP	mitochondrial Targeting Peptide

NCBI	National Center for Biotechnology Information
NF-H	Neurofilament Heavy subunit
NIH	National Institutes of Health
NIX	Nucleotide Identify X
nm	nanometre
NMD	Nonsense Mediated Decay
OPD	O-Phenylenediamine
ORF	Open Reading Frame
P21	Post-natal day 21
PBS	Phosphate Buffered Saline
PBS-T	Phosphate Buffered Saline-Tween®- 20
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PIX	Protein Identify X
PKR	Protein kinase, interferon-inducible double stranded RNA dependent
PTC	Premature Termination Codon
RC	Reliability Class
RIPA	Radioimmunoprecipitation
RNA	Ribonucleic acid
RNAi	RNA interference
rRNA	ribosomal RNA
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
SAGE	Serial Analysis of Gene Expression
SDS	Sodium Dodecyl Sulphate
siRNA	small interfering RNA
SMN	Survival Motor Neuron
SP	Signal Peptide
Taq	*Thermus aquaticus*
TBE	Tris-Borate-EDTA

TE	Tagging enzyme
TEMED	N,N,N',N'-Tetramethylethylenediamine
TNK2	tyrosine kinase, non-receptor, 2
trEMBL	translated EMBL
tRNA	transfer RNA
uORF	upstream Open Reading Frame
UTR	Untranslated region
UV	Ultra Violet
wst	wasted
ZFP259	Zinc Finger Protein 259

CHAPTER 1

INTRODUCTION

1. Introduction

1.1 Introduction

The aim of this thesis was to determine if a newly discovered gene contributed to the pathology of the mouse mutant wasted. In order to introduce this work, therefore I will first discuss the pathology of the wasted mouse before going on to address the nature of the genetic lesion responsible for this pathology. Finally I will discuss the reasons why I chose to examine an additional novel gene to assess its contribution, if any, to the phenotype of this mutant.

1.2 Pathology of the wasted mouse

In 1972 an autosomal recessive spontaneous mutation *wst* (wasted) arose in the inbred HRS/J colony maintained in the Mouse Mutant Stocks Centre at The Jackson Laboratory[1]. Mice homozygous for this mutation appeared normal directly after birth but at 21 days of age began to develop signs of neuromuscular degeneration *i.e.* wasting of muscles, tremor and uncoordinated body movements. The affected mice did not live beyond 30 days. In addition they also displayed sudden progressive hypoplasia of the spleen, thymus and lymph nodes, again beginning the third week of life. The spleen and thymus begin to atrophy and, by the time of death, the thymus in particular has almost disappeared. (C.M. Abbott personal communication)

1.2.1 Neuromuscular degeneration

Examining the neurological aspects of the wasted pathology in more detail, it was found that these mice displayed weight loss, tremors, incoordination and a decreased interest in their surrounding environment from post-natal day 21 (P21). Hind-limb paralysis followed, and by 28 to 31 days these mice were dead. The tremor observed involved all four limbs and was of high frequency but low amplitude. On a cellular level, these mice suffered from extensive vacuolar degeneration of the anterior horn cells of the spinal cord with less severe abnormalities in the motor nuclei of the brainstem including an abnormal accumulation of phosphorylated neurofilament heavy subunit (NF-H) in the perikarya of motor neurons[2].

1.2.2 Immune system abnormalities

In 1982 Shultz *et al*[1] described the marked lymphoid hypoplasia with significantly decreased spleen, thymus and lymph node to body weight ratios by 28 days of age that is characteristic of the *wst/wst* mouse. It was also noted that this hypoplasia was not restricted to a single major lymphocyte sub-population, and that *wst/wst* mice exhibited an impairment of delayed-type hypersensitivity (DTH) response to sheep red blood cells - *i.e.* significantly decreased DTH responses 4 days after priming compared to heterozygous controls. In addition, the authors noted a four-fold greater incidence of chromosome damage (breaks, gaps and fragments of chromosomes) in the bone marrow of *wst/wst* mice than in control littermates. When γ irradiated, 80% of metaphase *wst/wst* cells exhibited chromosomal damage compared to 30% of controls.

This paper proposed that mice homozygous for the *wst* mutation could be an animal model of ataxia telangiectasia (AT). This is an autosomal recessive disorder characterised by cerebellar ataxia, telangiectases, immune defects, and predisposition to malignancy. Chromosomal breakage is also a feature. AT cells are abnormally sensitive to killing by ionising radiation and abnormally resistant to inhibition of DNA synthesis by ionising radiation. Thus the initial hope that wasted mouse would be a good model for AT was a logical one.

However, this was found not to be the case, initially as result of mapping studies linking AT to chromosome 11q[3] (the *wst* mutation had been mapped to chromosome 2[4] in a region exhibiting conservation of synteny with human chromosome 20). This was then confirmed by the discovery of the ATM (Ataxia Telangiectasia Mutated) gene[5] and the wasted deletion[6].

Nevertheless, it was this hypothesis that biased much of the early research into the *wst* mutation. This meant that much of the discussion of the results found in these papers were concerned with whether or not the various observations made were similar to the AT phenotype – making for a large amount of published negative

results; *e.g.* The only result published in a paper by Nordeen *et al* in 1984[7] was that neither γ irradiation nor bleomycin-resistant component of DNA replication characteristic of AT fibroblasts was observed in cell lines from *wst/wst* mice. But not all previous work was so narrow in its viewpoint.

In 1986 Tezuka *et al*[8] noted the age-dependent increase in the frequency of spontaneous chromosome abnormalities in bone-marrow cells, and that in the *wst* lymphoid system the frequency of chromosome aberrations induced by γ irradiation reaches 13-fold that of normal mice by day 26. An absence of γ ray-induced mitotic delay was not apparent in *wst/wst* bone marrow, yet when the primer-activating enzyme activity was examined (a function of cellular DNA repair), there was no decrease in activity at 24 days. This suggested that the chromosomal aberrations seen were not as a result of dysregulation of the DNA repair function of the affected cells. In addition, the fact that γ rays could not induce a mitotic delay begins to point toward a potential deficit in the cell cycle checkpoint. This suggested that rather than being unable to repair damaged DNA, the *wst/wst* cells could not sense the damage that was occurring to their DNA and so continued to replicate and proliferate, resulting in an unusually large number of cells with chromosome abnormalities.

This conclusion was reinforced by work done by Inoue *et al*[9] in 1986 as this group looked at the levels of DNA synthesis following treatment of *wst/wst* spleen with bleomycin, 4NQO (4-nitroquinoline 1-oxide), γ irradiation and UV irradiation. It was discovered that following treatment with bleomycin, while control spleen cells' DNA synthesis decreased by 80%, that of *wst/wst* spleen cells decreased by only 30%. Similar results were found for 4NQO, and there was also a significant difference between controls and wasted mice for γ irradiation, but not as large a difference. UV radiation did not result in a difference between control and *wst/wst*. Bleomycin is a radiomimetic which exerts its effect by a sequence-specific degradation of DNA[10]. 4NQO (a UV mimetic) binds to DNA and produces adducts resulting in base substitution[11]. This is unusual as although 4NQO is primarily used as a UV mimetic, UV radiation itself did not distinguish *wst/wst* cells from wild type cells. However this could be because 4NQO does not produce a unique lesion as UV does but rather

multiple forms of DNA damage some of which could resemble γ irradiation. The DNA defects caused by bleomycin, 4NQO and γ irradiation should be devastating insults to the cell, and induce apoptosis/cell cycle arrest. This is not the case in *wst/wst* lymphoid cells, and so again we have evidence of the uncoupling of the DNA damage-sensing apparatus of the cell and the cell cycle checkpoint mediators.

Van Buul *et al* in 1987[12], and again in 1991[13], brought two new observations to light. Firstly in 1987 that *wst/wst* polychromatic erythrocytes exhibited an enhanced spontaneous rate of micronuclei formation – which reflects the prevalence of chromosome breakage. Secondly, in 1991 that translocation induction in the germ cells of *wst/wst* mice irradiated at 3-4 days old and examined at 21 days was normal. This implied two conclusions:

- The *wst* mutation is not a good model of AT.
- The problems inherent in the *wst/wst* ability to sense DNA damage exemplified by the lymphoid chromosomal aberrations, was not a universal phenomenon.

This last point is highly significant and is reinforced by the previous work showing that the abnormal response to radiation was not seen in *wst/wst* fibroblasts either[7]. Indeed, it suggested that the component of a checkpoint pathway affected by the *wst* mutation was a tissue-specific one.

A 1996 paper by Libertin *et al*[14] began to make interesting observations of *wst/wst* mice without reference to AT. One such observation included an increase in the mRNA expression from *Il6* (Interleukin 6), *Il1* (Interleukin 1 complex), *Crp* (C-reactive protein, petaxin related), *Jun* (Jun oncogene), and *Pdcd2*, (Programmed cell death 2) in *wst/wst* liver as compared to controls. It is worth noting that PDCD2, IL6 and JUN are involved in the induction of transcription associated with the onset of apoptosis. This group had previously documented decreased expression of interleukin 5 (thymus and spleen) and interleukin 2 receptor (thymus only) with increased expression of interleukin 2, γ interferon and transforming growth factor, beta 1[15].

When trying to interpret data of this nature, it is imperative to bear in mind the distinction between primary and secondary phenotypes. While there may indeed be various increases and decreases in expression of these cytokines in the liver, the animals are undergoing severe muscle wasting and neurological abnormalities at this stage. Therefore it is possible that these fluctuations in cytokine expression are simply reflecting the fact that these mice are stressed and/or are not feeding as normal rather than highlighting a specific cytokine pathway deficiency.

A more directed approach was taken by Woloschak *et al*[16] in studying the regulation of thymus PCNA (proliferating cell nuclear antigen) expression in wasted mice. This study was initiated following a 2D gel electrophoresis experiment where one protein in particular could be seen to be markedly decreased in *wst/wst* thymus in comparison to wild type and heterozygous controls. (In fact in total the authors claim that eight proteins were expressed at greater levels, and twenty-three at decreased levels in *wst/wst* mice relative to controls, but only one particular protein was studied further)
Following sequence analysis and inter-species comparisons, it was subsequently discovered that this protein was in fact the mouse homologue of PCNA (Proliferating Cell Nuclear Antigen). It was also noted that this decrease was not repeated in brain tissue. PCNA is a processivity factor for DNA polymerase δ with its expression associated with cell cycle progression. Given this and the fact that *wst/wst* lymphoid cells exhibit a high spontaneous rate of apoptosis (meaning the cells are not progressing through the cell cycle), some aspects of the *wst/wst* cell cycle were examined. It was found that while thymocytes responded similarly to controls with regard to concanavalin A activation, stimulation of some lymphoid cell sub-populations with mitogens, particularly in the spleen, caused an alteration in the fraction of cells in G_2 and S phase of the cell cycle.

Potter *et al*[17] in 1998 noted an increased level of induced and spontaneous apoptosis in thymocytes *in vivo* and *in vitro* while the brain, lung and gut showed no such increase. These results have been confirmed and expanded in our lab, with increased

apoptosis seen in splenocytes also (A.H. Wyllie and C.M. Abbott, personal communication).

While much of the previous research done into the immune system pathology of the wasted mouse points toward a deficiency in the maintenance of the correct cell cycle checkpoints, the nature of the cell cycle disturbance seen in this aspect of the wasted mouse phenotype remains unclear. In addition how this phenotype is connected to the neuromuscular deficiencies observed is also unknown.

1.3 The *wst/wst* genotype and *Eef1a2*

In 1998, using a positional cloning/positional candidate strategy, our research group discovered the genetic lesion responsible for the *wst* mutation[6]. Wasted mice exhibit a 15.8kb deletion on distal chromosome 2 that removes the promoter region and first non-coding exon of the mouse *Eef1a2* gene, abolishing transcription of this gene.

1.3.1 *Eef1a2*.

Eef1a2 was originally described by Ann et al[18] as a rat cDNA clone S1 which was discovered during the search for the Statin gene. It was found that the derived amino acid sequence of S1 shared over 92% homology with the human translation elongation factor eEF1α.

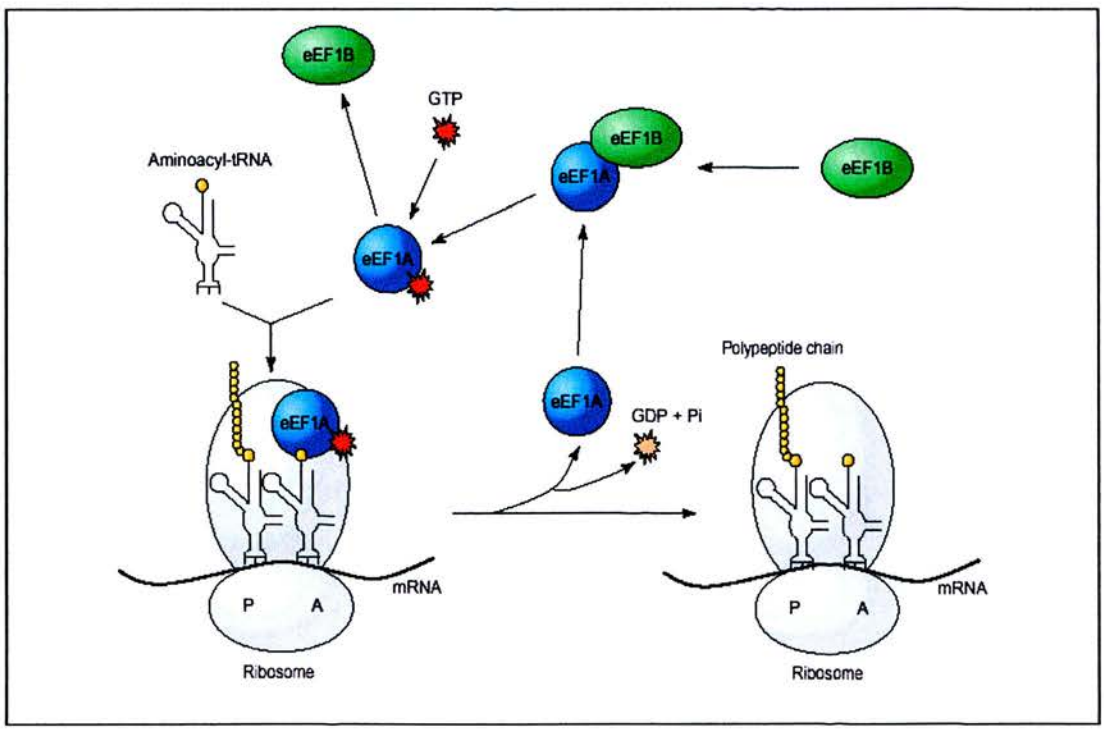


Figure 1.1 Schematic diagram of eEF1A's role in translation elongation (from ref [19])

eEF1α, when complexed with GTP, delivers the aminoacylated tRNA to the A-site of the ribosome. When appropriate codon–anticodon recognition occurs, GTP is hydrolysed and eEF1α–GDP is released from the ribosome. eEF1α then interacts with eEF1B, thereby promoting the exchange of the bound GDP for GTP to regenerate active eEF1α –GTP[19]

However the fact that the sequences of *S1* and EEF1α's 5' and 3' UTRs were less than 20% homologous implied that *S1* could be a separate distinct gene, regulated independently from EEF1α. This was also implied by the difference in temporal expression of mRNA from *S1* and *Eef1α* in the 3T3 mouse fibroblast cell line. While *S1* transcripts were found to be most abundant in the G0 phase of the cell cycle of this cell line but decreasing upon entering the G1 and S phases of the cell cycle, *Eef1α* transcripts showed no such changes in abundance upon progression through the cell cycle[18].

Again the separate nature of *Eef1α* and *S1* was highlighted when the same group found that *S1* was in fact expressed in a highly restricted manner in both mice and rats. Using RNase protection assays and Northern blots they found that the transcript of murine *S1* was present only in the brain, heart and muscle *i.e.* organs with a large proportion of terminally differentiated cells [20]. It soon became clear that *S1* was a *bona fide* gene similar to *Eef1α*, but regulated differently and the nomenclature was changed to reflect this with *Eef1a1* referring to the original *Eef1α* and *Eef1a2* referring to *S1*. In 1996 the loci of these two genes were discovered in humans with EEF1A1 found on chromosome 6q14 and EEF1A2 on 20q13.3[21]. Comparing mRNA expression from *Eef1a1* and *Eef1a2* in rabbit it was found that *Eef1a1* expression was not ubiquitous as previously thought, as it was not expressed in the skeletal muscle. Meanwhile *Eef1a2* was found to be expressed in the skeletal muscle, heart, brain and aorta[22]. In addition when the post-translational modifications of the two proteins were examined upon amino acid sequencing, it was found that while some remained the same (both of the glycerylphosphorylethanolamine modifications in eEF1A1 were found in eEF1A2) others differed (residues found to be dimethylated in eEF1A1 appeared to be trimethylated in eEF1A2).[22] With regard to the

functionality of these two proteins, however, both these proteins had indistinguishable activity in an *in vitro* translation system[22].

1.3.2 *Eef1a2* and wasted mice

The mouse *Eef1a2* gene was mapped by our group in 1998 and the wasted deletion uncovered[6]. Once *Eef1a2* was mapped close to *wst i.e.* the distal portion of mouse chromosome 2, it became clear that this gene was a good candidate. In comparing *wst/wst* and wild type mice it was found that PCRs designed to amplify exons I and II from *wst/wst* mice failed, while PCR products were obtained easily from wild type mice. This indicated the genetic lesion in wasted mice might be a deletion. Subsequently, through a process of screening an ES cell-derived BAC library, sequencing the genomic region surrounding *Eef1a2* and carefully designed PCR experiments, the boundaries of this deletion were characterised. One end of the deletion lay 206bp upstream of exon II of *Eef1a2*, with the other end lying within the gag gene of an IAP element upstream from the *Eef1a2* gene. Upon examining this deletion sequence there were no known or predicted genes present. It appeared therefore that the only gene directly affected by the deletion was *Eef1a2* which lost its promoter and first exon (although this exon is non-coding). To determine the effect of this insult to *Eef1a2* in *wst/wst* mice, the expression of the mRNA was examined and found to be completely abolished.

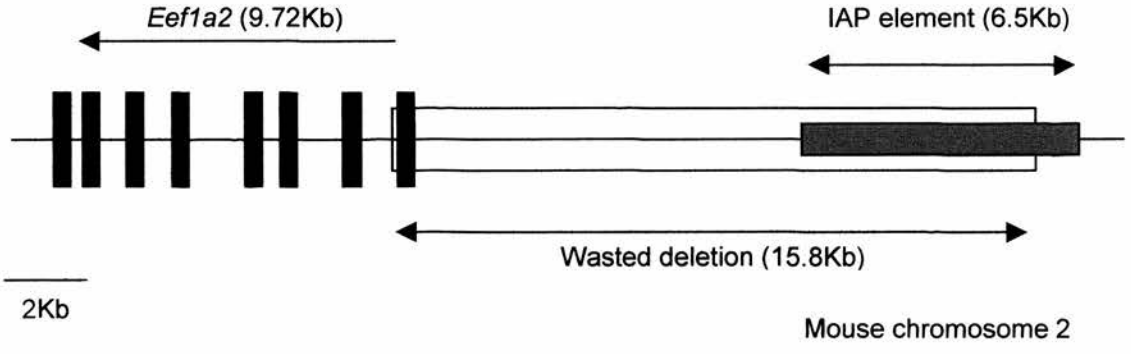


Figure 1.2 Diagram showing position of wasted deletion relative to *Eef1a2* locus

1.3.3 *Eef1a2* and the neuromuscular phenotype of wasted mice.

1.3.3.1 *Eef1a2* in skeletal muscle

When the expression pattern of *Eef1a1* and *Eef1a2* were examined more closely over time an interesting observation was made. While it had previously been reported that there was no *Eef1a1* expression in skeletal muscle in the rabbit[22] a more in-depth analysis found that skeletal muscle did express *Eef1a1* in the wild type embryo, but that this expression tailed off rapidly following birth with almost no *Eef1a1* expression detectable by P21. This expression profile was almost the opposite to that of *Eef1a2* whose expression levels increased dramatically following birth and begin to plateau at about P17 or P18[6]. These expression patterns are illustrated in the diagram below (figure 1.3).

It therefore appeared that, in skeletal muscle cells at least, eEF1A2 could "take over" from eEF1A1 *i.e.* eEF1A1 is not essential for living cells. If wasted mice have no *Eef1a2* expression it can be easily seen that by P21 these mice will have no functional eEF1A isoform in muscle cells and so there is no *de novo* translation of mRNA. It is therefore obvious that a lack of eEF1A2 could cause the muscle wasting and failure to gain weight seen in wasted mice from P21 onward.

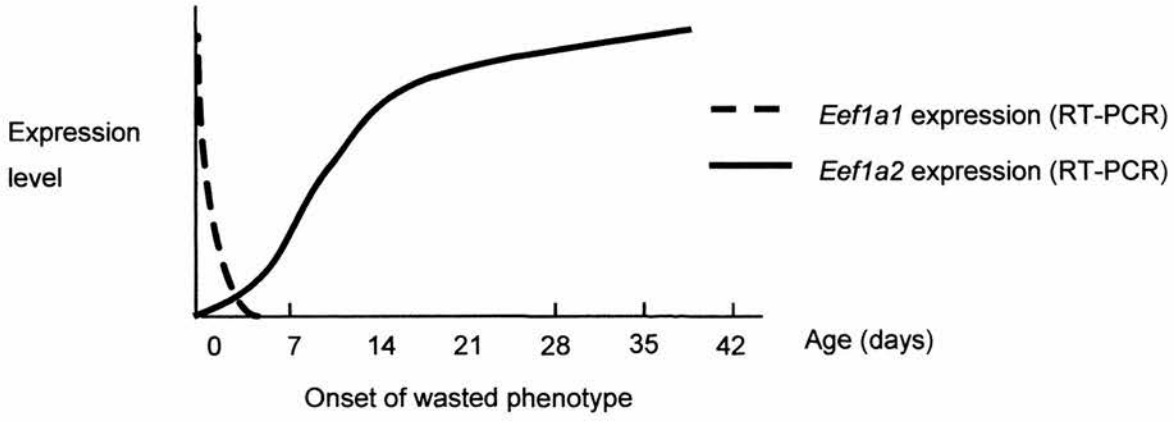


Figure 1.3 Diagram of developmental switching between *Eef1a* isoforms in skeletal muscle.

1.3.3.2. *Eef1a2* in the central nervous system

Both eEF1A1 and eEF1A2 are expressed in the central nervous system and so even with the loss of eEF1A2, there should still be a functional translation elongation factor present to carry out translation. When the expression of eEF1A1 and eEF1A2 was more closely examined in the central nervous system, however, it appeared that in the brain of adult wild type mice eEF1A2 is localised in neurons while eEF1A1 is found in non-neuronal cells. In neurons prior to P7 eEF1A1 is the major isoform but is later replaced by eEF1A2 which by P14 is the only isoform present[23]. This work was carried out largely in P20 mouse brain, but some recent evidence using specific eEF1A isoform antibodies in the spinal cord suggest that eEF1A1 is present in the nuclei of motor neurons, while eEF1A2 localises in the cytoplasm (H.J. Newbery personal communication). This would seem to contradict the previous study[23], but it can be seen that this paper only shows individual images of a eEF1A1-stained glial cell and a eEF1A2-stained neuron. Because these images are not double-labelled, it is unclear whether there would be any staining of a neuronal nucleus with the eEF1A1 specific antibody. In addition these studies were carried out at different sites, in the brain and spinal cord, and by different methods, immunohistochemistry and immunofluorescence. The different sensitivities of the methods used to detect eEF1A1 and eEF1A2 staining could be affecting the expression pattern observed or, as the pathology of the wasted mouse is seen largely in the spinal cord, it could well be that there are different expression patterns in these two components of the central nervous system.

What this work has shown is that while both isoforms of eEF1A are expressed in the central nervous system it appears that on a cellular and subcellular level the expression patterns of these isoforms diverge resulting in organelles and/or cell types which would be without any eEF1A isoform at all in a *wst/wst* mouse. This can be interpreted therefore as evidence that the loss of eEF1A2 could account for the neuromuscular pathology seen in the wasted mice.

In addition to its role in translation elongation, eEF1A1 has also been shown to exhibit other functions. It interacts with the transcription factor ZFP259 (Zinc Finger Protein 259), which in turn interacts with SMN (Survival Motor Neuron). Binding of both SMN and eEF1A1 to ZFP259 is required for their translocation from the cytoplasm to the nucleus. It is therefore possible that all three proteins are part of a complex. Given the similarity of eEF1A1 and eEF1A2, it is possible that eEF1A2 also interacts with these proteins. It is worth noting that SMN is deleted or mutated in >95% of cases in spinal muscular atrophy. This is a form of motor neuron disease that predominantly affects young children. Features of this disease include the loss of motor neurons of the anterior horn region of the spinal cord, the brain stem and the motor cortex and atrophy of the muscle fibres. It is possible therefore that the post-natal lack of both eEF1A1 and eEF1A2 in wasted muscle causes the downstream pathology through this ZFP259-SMN complex. However it is also possible that the motor neurons are severely affected by the lack of *de novo* protein synthesis because of their high metabolic requirement.

Further evidence that the loss of eEF1A2 expression could cause the neuromuscular deficiencies seen in wasted mice include a study where it was found that eEF1A2 could interact with the M_4 muscarinic acetylcholine receptor subtype. It had previously been shown that muscarinic activation increases dendritic translation[24] and although the subtype was not specified one can speculate that perhaps it is through eEF1A2 that this is accomplished. This suggests a specific role for eEF1A2 in translation at the dendrites of neurons.

Additional work has found that it is upon myogenic differentiation, the expression of eEF1A2 is activated[25].

These studies, in addition to elucidating how the loss of eEF1A2 expression could cause such damage to the wasted neuromuscular system, also begin to introduce the idea that eEF1A1 and eEF1A2 may have differing roles within the cell, and so even when eEF1A1 is present, this is not enough in cell types that usually express both, *i.e.* eEF1A2 has a role for which eEF1A1 cannot compensate. While it has been

shown *in vitro* translation systems that both eEF1A1 and eEF1A2 can function as translation elongation factors[22], there have been several studies that suggest that eEF1A1 plays a role in other systems in addition to translation.
One of the major functions eEF1A seems to play in the cell is with relation to the cytoskeleton. It has been found that eEF1A binds actin filaments and microtubules *in vivo* and *in vitro.* It binds and crosslinks actin in *Dictyostelium* amoebae[26] and *S. cerevisiae*[27]. There is also evidence that it can function in the bundling and severing of microtubules – although there has been some dispute as to whether eEF1A can sever microtubules[28, 29]. It has also been found to be an integral part of the mitotic spindle organising centres in sea urchins[30].

Some of the more interesting studies describe how, not only does eEF1A bind actin but also it appears to crosslink filament actin (F-actin) in particular, to form unique bundles[31]. Also when eEF1A is overexpressed in *S. cerevisiae* there are no effects on the efficiency or fidelity of translation, but *in vivo* there is an effect on the morphology of the cells. In particular there is an increase in the fraction of unbudded cells and the cells are larger and rounder in shape[27].

Why a translation elongation factor would have these functions is unclear but it could be surmised that they grew out of its original role as a translation factor. eEF1A, mRNA, ribosomes and actin can be co-localised in the cell[32] which suggests that a function of eEF1A, in addition to translating mRNA, has become to transport and anchor the mRNA to actin via its actin binding properties. It can be surmised from this that elongation factors' facility for binding and bundling actin as part of its translation role has had offshoot effects perhaps because of its abundance in the cell (eEF1A constitutes 1-2% of total protein in normal growing cells[33]). For example when *Dictyostelium discoideum* is stimulated by a chemical and begins to respond to that chemical with movement, this chemotaxis involves a sudden and dramatic increase in eEF1A -actin binding[26]. Elsewhere it has been found that in cell lines of higher metastatic potential there is a 30% reduction in the ratio of eEF1A to F-actin in the cytoskeleton which the authors suggest could allow for altered cytoskeletal organisation[34].

It is clear therefore that eEF1A does more than just make protein in a cell. But what does this information have to do with the loss of eEF1A2 in the central nervous system? It means that although we know that eEF1A2 can elongate polypeptide chains as eEF1A1 does, we are still unsure as to how much if any of the additional functions ascribed to eEF1α in the literature are also applicable to eEF1A2. In many of the studies I have discussed the work has been done in "eEF1α", with the authors unaware of the two isoforms. This can subtly change the conclusion one can draw from the experiments conducted. An example is the study on metastatic rat mammary adenocarcinomas. As I've discussed before, this group found that there is a 30% reduction in the ratio of eEF1A to F-actin in the cytoskeleton of these cell lines[34]. However this work was done using an antibody that would be predicted to recognise both eEF1A1 and eEF1A2 as it was designed against a peptide sequence that is identical in both of these proteins. Therefore the 30% reduction seen could actually be not only a reduction in total eEF1α but a switch from eEF1A1 to eEF1A2 for example. If these two isoforms then had different actin bundling properties or efficiencies it could mean that it is the switch in isoforms that causes the increased metastatic potential.
Therefore in the case of the brain and spinal cord – the loss of one isoform – despite the presence of the other – could still have wide-ranging effects on the cell.

In particular given the myocyte differentiation study[25] – it could be that one form is required for the growing and differentiating cell (eEF1A1) while the other is needed for the terminally differentiated cell with definitive architecture (eEF1A2). And that these differing roles are due to their having different actin or microtubule binding properties rather that any change in the translation elongation role of these proteins.

1.3.4 *Eef1a2* and the immune system.

A problem arises when addressing the immune system component of the wasted pathology with relation to the loss of *Eef1a2* expression. Several studies using a range of methods including specific Northern blots[20, 22] (with probes from the 3'UTR of *Eef1a1* and *Eef1a2* as these are less than 20% similar), RNase protection assays[20] and Western blots[23] with specific antibodies have concluded that there is no expression of *Eef1a2*/eEF1A2 in mouse[23], rat[20, 23], rabbit[22] or human[23] spleen. Given that this organ is one of the major sites of the immune system pathology, it is unclear how the loss of expression of *Eef1a2* in wasted mice could effect the changes seen in the spleen.

As discussed earlier the problems seen in the immune system such as the atrophy of the spleen and thymus, the increase in spontaneous chromosome aberrations, abnormal responses to DNA damaging agents, changes in rate of apoptosis appear to point toward some sort of dysregulation of the various checkpoints of the cell cycle in this system. Could it be possible that the loss of *Eef1a2* expression would cause these sorts of problems? There has been some evidence that eEF1A2 can play a role in the regulation of the growth and survival of the cell.

When *Eef1a1* is hypo- or hyper-expressed in Balb/c 3T3 cells there are subsequent changes in the apoptosis rate. Upon induction of serum starvation antisense *Eef1a1* was found to protect cells to some degree from apoptosis while *Eef1a1* over-expression results in a faster rate of cell death[35]. Hydrogen peroxide can be used to treat cultured cells in order to induce apoptosis by oxidative challenge. When this was carried out on H9c2 (a rat heart ventricle-derived clonogenic cell line) it was found there was a rapid increase in eEF1A protein levels which is thought to facilitate the apoptotic response to the oxidative stress [36] It is worth noting that the immunoblot conducted in this study was done with a "polyclonal eEF1α antibody". From the information given it is impossible to tell if this is an eEF1A1-specific antibody or one that would pick up both eEF1A isoforms. A 2002 study looked at the eEF1A isoforms in differentiating myocytes and found that while eEF1A2

expression is activated during differentiation, upon serum-starvation, eEF1A2 expression declines and eEF1A1 expression is turned on again in the dying myotubes[25]. This suggests that not only do these isoforms have an effect upon apoptosis but appear to have some sort of reciprocal relationship with regard to differentiation and apoptosis.

Given therefore that it has been shown that eEF1A1 and potentially eEF1A2 could have an effect on the growth and survival of cells it could be possible that the loss of eEF1A2 would have an adverse effect the apoptosis rate/cell cycle of a particular tissue. However this still does not explain how a tissue not expressing *Eef1a2* could be affected by its loss.

There was only one connection between eEF1A2 and the immune system which was revealed in an immunology paper which addressed T-cell tolerance to endogenous proteins expressed in a tissue specific manner[37]. To investigate why these proteins are not seen as non-self, the authors used eEF1A2 as a prime example of such a protein, as although it is highly similar to eEF1A1 it is only expressed in the skeletal muscle, heart and brain. During the course of this work, an RT-PCR experiment was used to look at expression of *Eef1a2* in the immune system. It was found that *Eef1a2* is expressed in bone marrow derived dendritic cells – which are antigen-presenting cells found in both the spleen and thymus.

Ablation of *Eef1a2* expression in dendritic cells, if true, could be seen as the proof that *Eef1a2* alone is responsible for the *wst/wst* phenotype. It is unclear, though, how the loss of a translation elongation factor could result in the type of chromosome damage observed in the lymphoid tissue especially as this damage was not confined to dendritic cells.

Dendritic cells are often called "professional" antigen presenting cells. They function by transporting antigens from the periphery into the inner tissues where lymphocytes reside. They do this by trapping, engulfing and digesting antigens of very diverse origin and then presenting these antigens to those lymphocytes that can take action

against the antigen sampled. On encountering antigens, a highly dynamic process is triggered in the dendritic cell, which allows it to migrate and change its phenotype and function. This process is referred to as maturation. Upon maturation, the dendritic cell is then able to convert the antigen into an efficacious immunogen and express the necessary cytokines and co-stimulatory molecules to galvanise the acquired immune system in action[38].

These cells are vital to the maintenance of a healthy and functioning immune system and it is possible to see that their malfunctioning through the loss of expression of a gene like *Eef1a2* would be devastating to this system. From what we know of eEF1A2 one could predict two possible outcomes for dendritic cells lacking a vital gene. Either the loss of eEF1A2 is so devastating that the dendritic cells are ablated; or, given its potential function in the control of the actin cytoskeleton, its loss results in a deficit in motility or the maturation process in the dendritic cell and thus hampers its proper functioning.

If we turn to the literature however to see if there is any precedent for similar insults to the dendritic cells causing the phenotype seen in wasted mice it seems unlikely that what we see in wasted immune system could be attributable solely to the loss of eEF1A2 in dendritic cells.
One study described the conditional ablation of dendritic cells in transgenic mice[39]. Using HIV regulatory sequences that control preferential expression in dendritic cells, mice were generated that expressed the HSV1-TK (Herpes Simplex Virus type 1 Thymidine Kinase) gene in dendritic cells. These mice were then treated with ganciclovir which is toxic in dividing HSV1-TK positive cells. This method allowed for the specific ablation of dendritic cells in the spleen and thymus such that 90% of splenic dendritic cells were depleted in a week. Following this treatment these mice exhibited major cortical atrophy of the thymus mainly due to the disappearance of CD4+CD8+ thymocytes. However there was no gross modification of spleen architecture[39].

In another study it has been found that patients with Wiskott-Aldrich syndrome, an X-linked recessive disorder, have intrinsic dendritic cell abnormalities[40]. This is due to the loss of function of the WASp protein due to mutations in the gene. WASp is an intracellular protein expressed exclusively in haematopoietic cells and is thought to regulate cytoskeletal architecture through the Rho family member TNK2[41]. Here, rather than loss of dendritic cells, the lack of functional WASp results in peripheral blood-derived dendritic cells that are rounded in appearance with actin abnormally distributed in perinuclear and subcortical regions. This results in dendritic cells that mature but cannot translocate. Given eEF1A2's potential for cytoskeleton interaction, this phenotype could be possible in dendritic cells lacking eEF1A2 – however the clinical aspects of Wiskott-Aldrich syndrome are immune deficiency, eczema and microthrombocytopaenia. Similarly young WASP-deficient mice showed mild thrombocytopaenia and lymphopaenia as well as defective T cell activation but no major atrophy or apoptotic defects[42].

These studies seemed to suggest that it was unlikely that the loss of eEF1A2 in wasted dendritic cells could explain the complete phenotype of the wasted immune system.

1.4 Identification of a novel gene

The mechanism by which the lack of a translation elongation factor in dendritic cells, which comprise only a small percentage of lymphocytes, could result in the type of chromosome damage observed in the lymphoid tissue of wasted mice was unclear.

It was important to remember therefore that the wasted genotype was not a targeted *Eef1a2* knockout but rather a classical mutation – a 15.8kb deletion. It therefore seemed important to question whether or not there were other genes affected by this deletion.

As addressed before there were no known or predicted genes within the wasted deletion and so it was possible to surmise that *Eef1a2* was the only gene directly affected in that its promoter and first exon were removed. However given the size of the deletion there could have been more subtle position effects on neighbouring genes. An enhancer or silencer element could have been removed by the deletion in wasted mice or the deletion could have changed the local chromatin environment such that the protein factors required for gene expression could no longer access such elements. This type of effect on a neighbouring gene could result in an increase/decrease in expression, ectopic expression or incorrect timing of expression of this gene. In this way even though the gene itself would not be mutated, there could be extensive pathology associated with its misexpression. There are many examples of human disease caused by position effects where the genetic lesion can be 10kb to 900kb away from the gene[43]

In order to address this possibility, it was important to take account of the genes surrounding the wasted deletion. The three known genes close to the wasted deletion did not present themselves as likely candidates for the immune system deficits in the wasted mouse. *Kcnq2* produces a voltage-gated potassium channel that is only expressed in the brain[44]. It is found to be mutated in Benign Familial Neonatal Convulsion (BFNC), a common epilepsy with autosomal dominant inheritance. In mice, *Kcnq2* knockouts die a few hours after birth of pulmonary atelectasis with no

evidence of brain abnormalities while *Kcnq2* heterozygotes display a hypersensitivity to a seizure-inducing agent pentylenetetrazole[45]. *Ptk6* produces a protein tyrosine kinase whose expression is confined to the skin[46] and gut[47]. There are no known phenotypes associated with loss of function of *Ptk6*.. *Srms* produces a Src-related kinase which is ubiquitously expressed but *Srms*$^{-/-}$ mice have no phenotype[48]. Meanwhile, closer than these three known genes, the careful sequencing of the region which allowed our group to elucidate the nature of the wasted genotype uncovered a predicted gene approx 10kb from the wasted deletion. This gene appeared to be completely novel.

Given what was known about *Eef1a2*, the function of dendritic cells and the features of the surrounding genes, I decided that the best course of action to elucidate the origins of the immune system phenotype was to use this novel gene as my prime candidate. The disadvantage of this course of action was that the gene and protein would have to be characterised and tools (such as antibodies) acquired to analyse its expression. However I felt that not only was this the best candidate but also its novelty meant that my characterisation of it would be important whether it turned out to play role in the wasted phenotype or not, while the other genes and proteins were well-characterised.

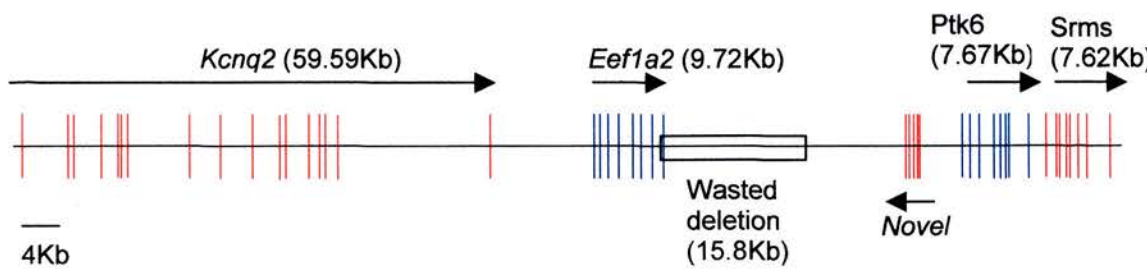


Figure 1.4 Genes surrounding the wasted deletion. While there are a number of known genes in the vicinity of the wasted deletion in addition to *Eef1a2*, none presented themselves as good candidates for the immune system phenotype of the wasted deletion. The gene closest to the wasted deletion other than *Eef1a2* was a novel undescribed gene.

1.5 Aims

The aim of this thesis therefore was to characterise this novel gene with a view to determining any role it played in the pathology of the wasted mouse. To do this I intended to:

- confirm its transcription to mRNA and translation to protein
- analyse its expression patterns in the body and in the cell
- assess its function through silencing the gene in mammalian cell culture

During the course of this work I aimed to have acquired the tools and knowledge to then examine the expression and function of this gene in wasted mice.

However, I could not neglect the potential that it was the loss of *Eef1a2* expression alone that was responsible for the whole of the wasted phenotype. To examine this possibility I hoped to analyse a set of transgenic animals in production by our group at the time of beginning this project that would determine the role of *Eef1a2* in wasted mice.

Chapters 3, 4 and 5 of this thesis describe the characterisation of the novel gene, while chapter 6 looks at the role of this novel gene and *Eef1a2* in the wasted mice.

CHAPTER 2

MATERIALS AND METHODS

2 Materials and Methods

2.1 Materials

2.1.1 General reagents

Unless otherwise indicated, general laboratory chemicals were obtained from Sigma. All PCR primers were from Sigma-Genosys or Invitrogen and cell culture media and supplements were obtained from Invitrogen.

2.1.2 Buffers and solutions

Recipes for those buffers and solutions not described in detail in the text are as follows:

10X Laemmli Running Buffer	250mM Tris HCl (pH 8.3), 1.9M Glycine 10% SDS
10X MOPS	200mM MOPS, 50mM sodium acetate 20mM EDTA, pH 7
20X SSC	3M NaCl, 300mM sodium citrate, pH7
2X Laemmli Loading Buffer	60mM Tris HCl (pH 6.8), 2% SDS, 0.1% bromophenol blue, 10% glycerol
Bejerrum & Schaefer-Nielsen Buffer	48mM Tris, 39mM Glycine, 20% MeOH 0.0375% SDS
ELISA OPD Buffer	0.493% w/v citric acid, 1.42% Na_2HPO_4, pH 5
ELISA Plate-coating Buffer	0.195% w/v Na_2CO_3, 0.293% w/v $NaHCO_3$ 0.02% w/v sodium azide, pH 9.6
ELISA Wash Buffer	0.8% w/v NaCl, 0.02% w/v KH_2PO_4, 0.114% w/v Na_2HPO_4, 0.02% w/v KCl, 0.5% Tween®-20, pH 7.4
Mowiol	7.5g of Mowiol 4-88® (Sigma) was mixed with 10ml glycerol in a 250ml flask. 25ml of dH_2O was added and the flask covered and left at room temperature overnight. 50ml of 0.2M Tris HCl (pH 8.5) was added and the

	mixture heated in a boiling water batch for 20 minutes. The mixture was allowed to cool before adding 1.75g of DABCO (Sigma), an anti-fade agent and then mixed until dissolved.
Northern Detection Buffer	0.1M Tris HCl, 0.1M NaCl pH 9.5
Northern Blocking Buffer	A 1:10 solution of the blocking solution provided by the manufacturer (Roche) in 0.1M maleic acid, 0.15M NaCl adjusted with NaOH (solid) to pH 7.5
Northern Washing Buffer	0.1M maleic acid, 0.15M NaCl, pH 7.5, 0.3% Tween®-20
PBS (Phosphate Buffered Saline)	1 PBS tablet (Sigma) was dissolved in 200ml of dH_2O and autoclaved
PBS-T (Phosphate Buffered Saline - Tween®-20)	As above but following the autoclave step 0.1% (v/v) of Tween®-20 (Sigma) was added.
Ponceau S	0.5% (w/v) Ponceau S, 2% (v/v) Acetic Acid
RIPA	150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50mM Tris HCl pH 8.0
RNA gel loading buffer	50% v/v formamide, 0.166% v/v formaldehyde, 10% v/v 10X MOPS, 10% glycerol, 0.02% 2.5% bromophenol blue, 0.005% (v/v) ethidium bromide, 0.109% (v/v) DEPC-treated dH_2O
TBE (Tris Borate-EDTA)	108g of Tris base, 55g of Boric acid and 9.3g of Na_4EDTA were added to 1L of dH_2O. pH 8.3 with no adjustment.

2.2 Methods

2.2.1 RT-PCR

2.2.1.1 RNA preparation from tissues or cell lines

RNA from both frozen tissue samples and cultured cell lines was extracted using the RNeasy Mini Kit (Qiagen) according to manufacturer's instructions.

2.2.1.2. DNase I treatment of RNA

RNA prepared as described above was subsequently treated with a DNase I enzyme to remove any contaminating genomic DNA. This was accomplished using the DNA-free Kit™ (Ambion) according manufacturer's instructions.

2.2.1.3 First strand cDNA synthesis

cDNA was prepared from isolated RNA by using the Retroscript® Kit (Ambion) according to manufacturer's instructions

2.2.1.4 PCR primers

Primers used to amplify portions of a range of genes were as follows:

Gene	Primer Name	Sequence
Clwd	TotalClwd-F	5'-ATCCACACTCCCGTTCATGA-3'
	TotalClwd-R	5'-TATGGGGTCAGGAAAGTTCA-3'
	Clwd A	5'-CGGTGCAGGAAAGCTACTAA-3'
	Clwd B	5'-AAACGAAGCCTGCAAGGAAG-3'
	Clwd C	5'-TCTCACCACGCAGCCGGACC-3'
	Clwd D	5'-TCCTGAAGCATGGCAGCCAT-3'
	Clwd E	5'-ATGGCTGCCATGCTTCAGGA-3'
	Clwd F	5'-CCAGGGTACTCGGCACTTCC-3'
Eef1a1	Eef1a1-F	5'-ATATTACCCCTAACACCTGG-3'
	Eef1a1-R	5'-CTGTGACAGATTTTTGGTCAAG-3'
Eef1a2	Eef1a2-F	5'-CCTCAGGAGGCTGCCCAG-3'
	Eef1a2-R	5'-ATGTCTCGCACGGCGAAGC-3'
Gapdh	Gapdh-F	5'-CATCACCATCTTCCAGGAGC-3'
	Gapdh-R	5'-ATGACCTTGCCCACAGCCTT-3'
Fn1	Fn1-F	5'-TTATGCTCAGAACCGGAACG-3'
	Fn1-R	5'-CGATGAGCTGTCTGGAGAAA-3'
Il1b	Il1b-F	5'-CTGTGACTCATGGGATGATG-3'
	Il1b-R	5'-GTATTGCTTGGGATCCACAC-3'
Ccl5	Ccl5-F	5'-TGCCCTCACCATCATCCTCA-3'
	Ccl5-R	5'-TTGGGTTTGCTGTGAGAGC-3'

These were designed across introns by and large in order to ensure that any genomic DNA contamination could be recognised easily.

2.2.1.5 PCR

The following components were added to sterile Eppendorf tubes on ice.

10X Sigma reaction buffer	3µl
10mM stock dNTPs	0.6µl
Forward primer (18pmol/µl)	1µl
Reverse primer (18pmol/µl)	1µl
dH_2O	21.1µl
Sigma *Taq* polymerase (5U/µl)	0.3µl
DNA template	3µl
Total	30µl

This reaction was then placed in a PTC-225® thermal cycler (MJ Research) set to the following programme of incubations; 94^{O}C for 3 minutes, then 94^{O}C for 30 seconds, 58^{O}C for 30 seconds, 72^{O}C for 30 seconds these three incubations were repeated for 29 cycles and finally 72^{O}C for 10 minutes.

2.2.1.6 Agarose gel electrophoresis

The PCR products were separated and assessed by gel electrophoresis. This was accomplished by making a 0.8% - 2 % (w/v) agarose solution in 0.5x TBE and heating until the agarose had dissolved. Once the solution had cooled slightly, but was still molten 0.003% ethidium bromide was added to the agarose solution and the mixture as poured into a casting tray with suitable sized comb to set. Once set, the tray was placed in a gel electrophoresis rig. The Ready Load^{TM} 1kb DNA ladder (Invitrogen) and samples of PCR products added to the wells in the agarose. The gel was then typically run at 100V for 1 hour. The PCR products could then be viewed in the gel under UV light.

2.2.2 Northern blotting

2.2.2.1 Production of probes

The DNA template was linearised at a restriction site downstream of the cloned insert in order to avoid transcription of undesirable sequences using a restriction enzyme that left 5' overhangs or blunt ends. The DNA was then purified by phenol/chloroform extraction, followed by ethanol precipitation.
The probes were labelled with DIG (digoxigenin) using the DIG Northern Starter Kit (Roche) according to the manufacturer's instructions.
The RNA produced during the reaction was then DNase I treated to remove template DNA; 2µl of RNase-free DNase I was added and the reaction was incubated for 15 minutes at 37^{O}C. The reaction was stopped by adding 2µl of 0.2M EDTA (pH 8) and a small sample of RNA product run on an agarose gel to check the size and purity of the probe.
The efficiency of labelling was determined by carrying out serial dilutions of the DIG-labelled probe and then applying the dilutions to nylon membrane. This was also done with the supplied DIG-labelled control of known concentration.

2.2.2.2. Phenol/Chloroform extraction

An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the nucleic acid sample in a polypropylene tube. The contents of the tube were mixed well by gentle inversion, until an emulsion formed. The organic and aqueous phases were then separated by centrifugation at 12,000g for 5 minutes. The aqueous phase (upper layer) was then carefully removed with a large-bore pipette, and the interface and organic phase discarded. The phenol:chloroform:isoamyl alcohol extraction was repeated until no protein (cloudy, white precipitate) was visible at the interface. To remove traces of phenol, the nucleic acid solution was then extracted with an equal volume of chloroform:isoamyl alcohol (24:1) in exactly the same manner.

2.2.2.3 Ethanol precipitation of DNA

The volume of DNA to be precipitated was measured, and the salt concentration adjusted by adding 1/4 volume of 10M ammonium acetate (final concentration 2.5M). After mixing well, 2 volumes of ice-cold 100% ethanol were added and the solution was mixed again. The DNA was then allowed to precipitate at on ice for 30 minutes. It was then pelleted by centrifugation at 12,000g for 15 minutes. The DNA was then washed twice in 70% ethanol and allowed to air dry. Finally the DNA was resuspended in an appropriate volume of dH_2O.

2.2.2.4 Ethanol precipitation of RNA

To 20µl of RNA, 2.5µl 4M LiCl and 75µl EtOH were added and incubated at $-20^{O}C$ for 2 hrs. This was then centrifuged at 16000g for 10 minutes. The solution was then washed in 70% EtOH (in DEPC dH_2O) and allowed to air dry. The pellet was then resuspended in 100µl DEPC dH_2O and 1µl RNase inhibitor and incubated for 1 hour on ice and then vortexed and incubated at $-20^{O}C$ overnight as this aids resuspension.

2.2.2.5 RNA gel electrophoresis

First the gel rig was sterilised with 10% SDS in DEPC-treated water and 100% ethanol. The gel was then prepared as follows:
3.6g of agarose were added to 255ml of DEPC-treated water and the mixture heated until the agarose had melted. This was then allowed to cool to approximately $60^{O}C$ before adding 28.38ml of 10X MOPS and 16.2ml of 37% formaldehyde in a fume hood. This gel was then poured and allowed to set for at least 30 minutes at room temperature. 20µl of loading buffer was added to the RNA samples and then incubated at $65^{O}C$ for 10 minutes and $4^{O}C$ for 1 minute. The gel was pre-run at 30V in 1X MOPS buffer. The RNA samples were then loaded and the gel run at 125V for 1-4 hours.

2.2.2.6 Northern blotting

The RNA gel was trimmed and photographed with a ruler before incubating it in DEPC-treated water for 15 minutes agitating at room temperature and then in 20X SSC for 15 minutes agitating at room temperature. The RNA gel was then placed on a wick of filter paper soaked in a bath of 20X SSC, nylon membrane placed on top then three sheets of filter paper on top of that and finally a wad of paper towels pressed down by a suitable weight to aid blotting. This was blotted overnight. The membrane was then removed from the blotting apparatus, placed on 2X SSC-soaked 3MM paper and UV cross-linked. The membrane was then rinsed in DEPC-treated water and allowed to air dry.

2.2.2.7 Hybridisation

The membrane was pre-hybridised in 20ml of DIG Easy Hyb solution (Roche) in a roller bottle at 68^{O}C for 1 hour. 800ng of RNA probe was then denatured by boiling for 5 minutes and chilling on ice. The probe was then added to 8ml of fresh DIG Easy hybridisation fluid and the pre-hybridisation fluid replace with the hybridisation fluid containing the probe. The membrane was incubated in this at 68^{O}C overnight.

2.2.2.8 Stringency washes and signal detection

The membrane was removed from the hybridisation fluid and washed in 2X SSC/0.1% SDS twice for 5 minutes each at room temperature and then in 0.1X SSC/0.1% SDS twice for 15 minutes each at 68^{O}C. The membrane was then incubated in washing buffer for 5 minutes and then blocking buffer for 1 hour. The anti-DIG antibody (Roche) was added to Northern blocking buffer and the membrane incubated in that for 30 minutes. The membrane was washed three times for 15 minutes each in Northern washing buffer. The membrane was then incubated in Northern detection buffer for 5 minutes and then incubated in CPD-*Star* (Roche) for 5 minutes. The membrane was then sealed in a plastic bag with the CPD-*Star*, exposed to X-ray film for 30 minutes and then the film developed.

2.2.3 Preparation and purification of anti-peptide antibody

2.2.3.1 Design of peptide to be used as an immunogen

The PIX (Protein Identify X) suite of programmes at the HGMP (Human Genome Mapping Project - www.hgmp.mrc.ac.uk) includes a programme that predicts the antigenicity of portions of the protein under investigation. This was used to design the peptide sequence that would be chemically synthesised and conjugated to KLH (Keyhole Limpet Haemocyanin) by Graham Bloomberg at the University of Bristol and subsequently used as an immunogen in sheep at Diagnostics Scotland. The peptide sequence chosen was IPSSGSLVATHDY.

2.2.3.2 Confirmation of antibody specificity by ELISA

Peptide was dissolved in dH_2O to a final concentration of 5 mg/ml. This was then used to make serial dilutions of peptide; 5μg/ml, 1 μg/ml, 0.5 μg/ml, 0.1 μg/ml, 10ng/ml and 1ng/ml. 200μl of each dilution were added to 12 wells of an ELISA plate (Nunc). This was then incubated overnight at $4^{O}C$. The wells were then washed three times with ELISA wash buffer before being blocked by (freshly prepared) 1% BSA in ELISA wash buffer and incubated for 1 hr at room temperature. The wells were then washed three times. The primary anti-peptide antibody was then diluted to 1:100, 1:1000, 1:10000 and 1:50000 in ELISA wash buffer and added to the wells for 1 hour at room temperature. The wells were again washed three times. The secondary antibody was then added, diluted in ELISA wash buffer (1:1000) and incubated at room temperature for 1hr. The wells were again washed three times. One 10mg tablet of OPD substrate (Sigma) was then added to 25ml of OPD buffer with 10μl of 30% hydrogen peroxide and added to each well. The absorbance of the reaction was measured by spectrophotometer at 450nm.

2.2.3.3 Ammonium sulphate precipitation of anti-Clwd antibody

The volume of the anti-serum was determined before centrifuging it at 3000g for 30 minutes. While the anti-serum was stirring gently 0.5 volumes of saturated ammonium sulphate solution was added slowly in a drop wise fashion. Once added the solution was transferred to $4^{O}C$ and allowed to stir for 6 – 16 hours. The solution was then centrifuged at 3000g for 30 minutes and the supernatant recovered. The addition of 0.5 volumes of saturated ammonium sulphate to the supernatant was repeated as before and allowed to stir again for 6-16 hours at $4^{O}C$. This was then centrifuged at 3000g for 30 minutes, following which the supernatant was removed and discarded and the pellet resuspended in 0.3-0.5 volumes of PBS, avoiding frothing. This was then dialysed versus three changes of PBS overnight

2.2.3.4 n-Octanoic acid precipitation of anti-Clwd antibody

The volume of the antiserum to be fractionated was measures and 2 volumes of 0.06M acetate buffer pH4.0 was added and allowed to stir. For each 100ml of anti-serum, 7.5ml of n-octanoic acid was added slowly drop wise and allowed to stir for a further 30 minutes following the addition. The precipitate formed was removed by centrifuging the mixture at 2000g for 20 minutes at room temperature. The supernatant was recovered and filtered before dialysis against 0.154M sodium chloride to remove the n-octanoic acid.

2.2.3.5 Immunoaffinity purification of anti-Clwd antibody

The ammonium sulphate or n-octanoic acid precipitated anti-serum was further purified by running through an immunoaffinity column. These were obtained from Pierce Biotechnology as a SulfoLink Kit with the purification being carried out as per manufacturer's instructions.

2.2.4 Western Blotting

2.2.4.1 Preparation of protein extracts

In order to extract protein from tissue samples, the tissue was homogenised in 0.32M sucrose, with Complete Protease Inhibitor Cocktail (Roche) added to protect against extract degradation. For cell lines, the cell pellets were resuspended and vortexed in RIPA buffer also supplemented with protease inhibitors

2.2.4.2 Quantitation of protein

This was carried out using the Bio-Rad *DC* Protein Assay Kit following manufacturer's instructions

2.2.4.3 Preparation of SDS-PAGE gel

The casting plates from Bio-Rad Mini PROTEAN 3 mini-gel set were assembled. The 12% separating SDS-PAGE gel was then prepared as follows:

30% acrylamide	12ml
1.5M Tris HCl pH 8.8	7.5ml
dH_2O	9.9ml
10% SDS	0.3ml
TEMED	0.012ml
10% APS	0.3ml

This was then poured between the plates leaving a 2cm space at the top for the stacking gel. The separating gel was overlaid with isopropanol to avoid bubble-formation and allowed to set for 30-45 minutes.

The isopropanol was decanted and the stacking gel prepared as follows:

30% acrylamide	1.45ml
1.5M Tris HCl pH 6.8	2.5ml
dH_2O	5.95ml
10% SDS	0.1ml
TEMED	0.005ml
10% APS	0.125ml

The stacking gel was then poured between the plates to the top and the combs inserted at a 45^O angle to avoid trapping air and pressed down firmly. This was then allowed to set for 15-30 minutes.

The plates were then placed into the gel tank apparatus, the inner chamber between plates filled with 1X Laemmli running buffer and the outer chamber half-filled. The combs were then removed and any acrylamide present in the wells was washed out. The protein extract samples were then treated with 1µl 1M DTT/10µl of sample and equal volume of 2X Laemmli loading buffer. This was boiled for 5 minutes and then loaded into the wells of the SDS-PAGE gel. A set of Pre-stained SDS PAGE Standards, Broad Range (Bio-Rad) was also run to assess protein size. The gel was run at 100V as the dye front progressed through stacking gel and increased to 150V once dye front had entered separating gel.

2.2.4.4 Western blotting

Filter paper, Hybond^{TM}-P membrane (Amersham Pharmacia Biotech) and the SDS-PAGE gel were soaked in Bejerrum Schafer-Nielsen transfer buffer. The filter paper then Hybond^{TM}-P membrane then the gel and then another piece of filter paper are placed on the semi-dry blotter. The gel was blotted at 15V for 1-2 hrs. The membrane was then stained in Ponceau S for 2 minutes, with the excess stain washed off with dH_2O in order to visualise the transferred proteins.

2.2.4.5 Detection of signal

The non-specific sites on the protein-containing membrane were blocked by agitating the membrane in 5% powdered milk in PBS-T overnight at $4^{O}C$. The membrane was then washed twice for 5 minutes each in PBS. The primary antibody was added to a fresh milk solution and used to incubate membrane for 2 hours agitating at room temperature. The membrane was washed 4 times for 5 minutes each in PBS-T. This was repeated with secondary antibody for 1 hour agitating at room temperature and then washed in PBS. Finally the membrane was either incubated in ECL™ solution (Amersham Pharmacia Biotech) for 1 minute, exposed to X-ray film and developed, or incubated in DAB (Sigma) for several minutes until the desired depth of colour change had taken place and the membrane was washed in dH_2O to stop the reaction

2.2.5 Maintenance of cell lines

2.2.5.1 Thawing of frozen vials of cells

Vials were removed from liquid nitrogen and carefully placed in a 37^{O}C water bath to thaw – taking care in case any liquid nitrogen had penetrated the vial as this can cause the screw cap to explode off when vial is heated. Once thawed, the cells were added to 15ml of medium in a T75 flask that had been at 37^{O}C and 5% CO_2 for at least 15 minutes. The flask was then returned to the incubator at 37^{O}C and 5% CO_2.

2.2.5.2 Subculturing of cells

Cells were passaged by first aspirating the growth medium from the flask/plate. 3ml of trypsin:versene (1:1) was added to wash out any remaining medium and then aspirated. A further 3ml of trypsin:versene was added and the flask placed at 37^{O}C for approximately 3-5 minutes until the cells had been detached from the plastic (monitored by microscope). 7ml of medium was then added and the cells aliquoted to the desired dilution in fresh flasks of medium. 15ml of growth medium was used in T75 flasks/ 0.5ml in wells of 24-well plates.

2.2.5.3 Counting cells using a haemocytometer

The cells were trysinised and then 100µl of the cell suspension is added to 100µl of Trypan Blue stain (Invitrogen). This mixture was then pipetted between the coverslip and slide of the haemocytometer and examined under the microscope. The number of cells (X) within the central square of the haemocytometer denotes the concentration of cells *i.e.* X x 10^4/0.5ml. Any cells stained blue were dead.

2.2.5.4 Cryopreservation of cell lines

Cells were trypsinised, counted and then resuspended drop wise in medium supplemented with 5% DMSO to a final concentration of 4×10^4 cells/ml and placed in screw cap cryopreservation vials on ice. The vials were removed to a $-70^{O}C$ freezer overnight and placed in liquid nitrogen the following day.

2.2.6 RNAi

2.2.6.1 Design of siRNA duplexes

siRNA duplexes specific for Clwd were designed exactly as described in ref [49]. Others described were obtained as library siRNAs available from Qiagen

siRNA	Target sequence
$Clwd^{+89}$	5'-AAGTGCCGAGTACCCTGGGGA-3'
$Clwd^{+183}$	5'-AAATCCACACTCCCGTTCATG-3'
Control (non-silencing) siRNA	5'-AATTCTCCGAACGTGTCACGT-3'
Vinculin	5'-AAGCGGATTAGAACCAATCTC-3'

2.2.6.2 Transfection of cell line

Cell lines were transfected with siRNA using two protocols. The first was exactly as described by ref [49] using Oligofectamine™ (Invitrogen), the second exactly as described by the manufacturer's instructions of the RNAi Starter Kit (Qiagen). The cells were examined for phenotype one, two, three and four days following transfection.

2.2.7 Immunofluorescent staining of cells

2.2.7.1 Fixing of cells

Cells were grown in 24 well plates on 13mm diameter coverslips (BDH). They were fixed by aspirating off the growth medium, washing briefly in PBS and then incubating in 0.2ml of an ice-cold 1:1 methanol:acetone mix at -20^{O}C for 5 minutes. The methanol/acetone was then aspirated and the cells washed in PBS three times before proceeding with staining.

2.2.7.2 Immunostaining and counterstaining of cells

Once fixed, the cells were washed twice with PBS and then three times with PBS-T. The cells were then incubated in a blocking solution (PBS-T/10% donkey serum/3% BSA) for 1 hour before being incubated in the primary antibody diluted in the above blocking solution for 1 hour. The cells were then washed three times in PBS-T before being incubated in the appropriate secondary antibody, again diluted in the blocking solution described, for 30 minutes. Following this, the cells were again washed three times in PBS before the coverslip was removed from the well and mounted (cell-side down) on SuperFrost® Plus microscope slide (Menzel Glaser) using Mowiol® (Sigma) with a non-fade reagent added. The slides were allowed to air dry for 30 minutes before being wrapped in tin-foil and stored overnight at 4^{O}C. The cells were visualised the following day on an Axioskop 2 fluorescent microscope (Zeiss) and using Smart Capture software

2.2.8 Analysis of mouse stocks

2.2.8.1 Mouse strain

The wasted mutation arose spontaneously in the inbred HRS/J colony at the Mouse Mutant Stocks Center of the Jackson Laboratory. The mutation was then crossed to a segregating C3HeB/FeJ x C57BL/6J background.

2.2.8.2 Mouse husbandry

Mice were housed at the MRC Evans Building Transgenic Unit at the Western General Hospital. They were fed on a standard chow diet. All mice euthanised were done so by means of a carbon dioxide chamber.

2.2.8.3 Preparation of DNA from ear notches

For the purposes of identification and genotyping, mice were ear notched. The ear notches were boiled in 0.6ml 50mM NaOH for 10 minutes, vortexed, neutralised with 50µl Tris HCl pH 8.0 and centrifuged for 6 minutes. The supernatant was recovered and 1-3µl used in the genotyping PCR.

2.2.8.4 Genotyping of mice

Mice were genotyped by PCR in order to ascertain the presence of the wst allele, the wild type allele and the BAC transgene or the PAC transgene

The PCR primers used were as follows.

To detect:	Primer Name	Sequence
BAC	B1162	5'-GAGACGTTGATCGGCACG-3'
	B1163	5'-CATGTTTGACAGCTTATCATCG-3'
PAC	pCYPAC2F	5'-CACTTACTCTCACTTCGC-3'
	pCYPAC2R	5'-CCATTCCGACAGCATCGCC-3'
Wild type allele	P2F	5'-TAGTGGCTCCTTGGAACAG-3'
	P2R	5'-CTACTCTCCCTGAATGCCTT-3'
Wst allele	Wstspan F	5'-ATAAGCTCCCCAATGGTAGAGAA-3'
	Wstspan R	5'-CGCGCCATTCTTGTATTGTT-3'

2.A Materials and methods appendix

2.A.1 SAGE (Serial Analysis of Gene Expression)

This section describes the method for gathering SAGE data. This data was queried in chapter 3 to gain information on Clwd expression.

Serial Analysis of Gene Expression (SAGE) is a high throughput technique for measuring gene expression profiles. Short cDNA tags (10-14bp) corresponding to unique transcripts are isolated and concatenated into single molecules, allowing for the simultaneous sequencing of thousands of tags. The relative tag abundance estimated in the sequencing step reflects the transcription patterns of the corresponding genes, allowing differential analysis of expression levels. How these tags are acquired is explained below.

mRNA is harvested from the tissue and captured on oligo-dT beads. Double stranded cDNA is then synthesized from the mRNA of the tissue or cells. This cDNA is then digested with a frequently cutting restriction enzyme, the anchoring enzyme (AE), with a 4-bp recognition site (usually NlaIII). A ligation reaction is performed which attaches a linker to the 3' overhang of the digested cDNA. From 5' to 3' the linker contains a primer sequence of either type A or B, a recognition site for a type IIS restriction endonucleases (the tagging enzyme, TE), and a recognition site for the AE. The nature of the TE will determine the length of the tag sequences obtained. Short (10 base) nucleotide sequences are often sufficiently specific to discriminate different transcripts from each other. The reaction is digested with the TE, which cleaves DNA at a position 14-18 base pair (dependent on the enzyme chosen) from its recognition site. This releases the linker along with short tags from a defined position within each of the mRNA transcripts, namely 3' of the AE restriction site closest to the Poly-A site. Pools of tag-linker oligonucleotides are ligated and subjected to a PCR reaction. The amplified products contain two tags (ditags) bordered by linker DNA. The linker DNA is removed from the ditags by a AE-digest. The obtained fragments are ligated together to form a concatemer of tags

which are cloned into a standard plasmid vector and sequenced. Concatenated arrays are combined to form SAGE "libraries". The number of times a particular tag is detected in a library gives a digital measure of the abundance of its associated mRNA and, hence, provides a quantitative measure of gene expression.

2.A.2 Useful websites

Site	URL
Target P	http://www.cbs.dtu.dk/services/
NetOGlyc	http://www.cbs.dtu.dk/services/
YinOYang	http://www.cbs.dtu.dk/services/
PredictProtein	http://cubic.bioc.columbia.edu/predictprotein/
REP	http://www.embl-heidelberg.de/~andrade/papers/rep/search.html
Expasy	www.expasy.ch
The siRNA User Guide	http://www.rockefeller.edu/labheads/tuschl/sirna.html
HGMP	www.hgmp.mrc.ac.uk
Ensembl	www.ensembl.org
The Institute for Genomic Research	www.tigr.org
Neural Network Promoter Predictor	http://www.fruitfly.org/seq_tools/promoter.html
Dragon Promoter Finder	http://research.i2r.a-star.edu.sg/promoter/promoter1_4/DPF.htm

2.A.3 Matched tumour/normal expression array patient data

Location	Tissue	Patient	Additional details	ICD Morphology of Neoplasms	ICD Neoplasm Classification
1A/1B	kidney renal cell carcinoma and normal tissue	66-yr-old male	right kidney; clear cell; grade II–III, partially cystic (about 4.5 cm in diameter in upper pole); no evidence of tumour in vessels, ureter & capsule Metastasis: none seen	not available	not available
2A/2B	kidney transitional cell carcinoma and normal tissue	73-yr-old female	right kidney; poorly differentiated with squamous differentiation (grade III–IV); extensive invasion seen in kidney, including renal tubules, lymphovascular areas & perinephric fat; no evidence of tumour in ureter, artery & vein Metastasis: not available	not available	not available
3A/3B	kidney renal cell carcinoma and normal tissue	44-yr-old female Caucasian	Metastasis: none seen	M8312/3	189.0
4A/4B	kidney renal cell carcinoma and normal tissue	57-yr-old male	left kidney; grade III–IV; 8.0 cm; extension through renal capsule; invasion seen in renal vein & hilar structures Metastasis: Metastases to bone	not available	not available
5A/5B	kidney renal oncocytoma and normal tissue	73-yr-old male	right kidney; 13.5 cm; history of haematuria Metastasis: not available	not available	not available
6A/6B	kidney renal cell carcinoma and normal tissue	72-yr-old female	right kidney; clear cell type; 5.0 cm; invasion seen in renal vein lumen in renal sinus; no tumour seen in ureter and vascular surgical margins Metastasis: none seen	not available	not available
7A/7B	kidney renal cell carcinoma	62-yr-old male	right kidney; clear cell type grade II/III; 8	not available	not available

	and normal tissue	Caucasian	cm; invasion seen in capsule, but confined by Gerota's fascia; no invasion seen in renal vein, ureter, and adrenal gland Metastasis: none seen		
8A/8B	kidney renal cell carcinoma and normal tissue	43-yr-old male Caucasian	right kidney; clear cell type grade III; 13 cm; extension into, but not through capsule; extensive tumour necrosis; no invasion seen in renal vein Metastasis: metastases to lymph node, adrenal gland & lung	not available	not available
9A/9B	kidney renal cell carcinoma and normal tissue	50-yr-old female	left kidney; clear cell type grade II/IV; 15 cm; partially cystic; no invasion seen in capsule; invasion seen in renal vein Metastasis: none seen	not available	not available
9G/9H	kidney renal cell carcinoma and normal tissue	57-yr-old female Caucasian	Metastasis: none seen	M8312/3	189.0
10A/10B	kidney renal cell carcinoma and normal tissue	59-yr-old male	right kidney; clear cell type grade IV; 13 x 9 cm; invasion seen in capsule, but confined by Gerota's fascia; no invasion seen in renal vein and hylar vessels; extension of tumour thrombus into inferior vena cava Metastasis: metastases to lymph nodes	not available	not available
11A/11B	kidney renal cell carcinoma and normal tissue	53-yr-old male	right kidney; clear cell type grade II/IV; 5 cm; no invasion seen in capsule and renal vein; neoplasm composed of cells arranged in an alveolar pattern; individual cells have moderately pleomorphic nuclei and abundant clear cytoplasm Metastasis: not available	not available	not available
12A/12B	kidney renal cell carcinoma and normal tissue	47-yr-old male Caucasian	Metastasis: none seen	M8312/3	189.0

13A/13B	kidney renal cell carcinoma and normal tissue	56-yr-old male Caucasian	Metastasis: none seen	M8312/3	189.0
14A/14B	kidney renal cell carcinoma and normal tissue	a 60-yr-old female Caucasian	Metastasis: none seen	M8312/3	189.0
1D/1E	breast infiltrating ductal carcinoma and normal tissue	48-yr-old	left breast; poorly differentiated; high nuclear grade; carcinoma shows squamous differentiation; minimal comedo-type DCIS with high nuclear grade; microcalcification seen in association with DCIS Metastasis: metastases to lymph nodes	not available	not available
2D/2E	breast infiltrative ductal carcinoma and normal tissue	59-yr-old	Metastasis: not available	not available	not available
3D/3E	breast infiltrating duct carcinoma and normal tissue	45-yr-old Caucasian	Metastasis: metastases to lymph nodes	M8500/3	174.9
4D/4E	breast medullary carcinoma and normal tissue	47-yr-old Caucasian	Metastasis: none seen	M8510/3	174.9
5D/5E	breast infiltrative lobular carcinoma and normal tissue	60-yr-old Caucasian	Metastasis: metastases to lymph nodes	M8500/3	174.9
6D/6E	breast infiltrative ductal carcinoma and normal tissue	50-yr-old	Metastasis: not available	not available	not available
7D/7E	breast infiltrating duct carcinoma and normal tissue	71-yr-old Caucasian	Metastasis: metastases to lymph nodes	M8500/3	174.9
8D/8E	breast mucinous adenocarcinoma and normal tissue	53-yr-old Caucasian	Metastasis: metastases seen in adipose tissue	M8480/3	174.9
9D/9E	breast infiltrating duct carcinoma and normal tissue	42-yr-old Caucasian	Metastasis: metastases seen in lymph nodes	M8500/3	174.9
11D/11E	prostate adenocarcinoma and normal tissue	65-yr-old	Gleason's grade 3 + 4 = 7; extension through capsule Metastasis: none seen	not available	not available
12D/12E	prostate adenocarcinoma and normal tissue	67-yr-old	Gleason's grade 3 + 5 = 8; focal extension through capsule Metastasis: none seen	not available	not available

13D/13E	prostate adenocarcinoma and normal tissue	68-yr-old Caucasian	Gleason's grade 3 + 4 = 7; 4.0 cm; extension through capsule; perineural invasion seen; no invasion seen in seminal vesicles or urethra; PSA of 29 Metastasis: none seen	not available	not available
1G/1H	uterus endometrioid adenocarcinoma and normal tissue	73-yr-old	moderately differentiated (grade II–III); extensive adenomyosis (with focal involvement by carcinoma); benign paraovarian cysts; leiomyoma; no evidence of tumour in anterior vagina; post menopausal Metastasis: not available	not available	not available
2G/2H	uterus keratinizing squamous cell carcinoma NOS and normal tissue	46-yr-old Caucasian	Metastasis: none seen	M8071/3	180.9
3G/3H	uterus benign tumour; menorrhagia & fibroids and normal tissue	55-yr-old	submucosal, subserosal & intramural leiomyomata; inactive endometrium with chronic inflammation & focal minimal; bilateral pyosalpinx & acute salpingitis Metastasis: not available	not available	not available
4G/4H	uterus benign tumour and normal tissue	76-yr-old	total vaginal prolapse (cervical hyperparakeratosis); focal cervical squamous metaplasia with dilated endocervical glands; atrophic endometrium; focally hyalinized leiomyomata Metastasis: not available	not available	not available
5G/5H	uterus benign tumour and normal tissue	55-yr-old	menorrhagia & fibroids; submucosal, subserosal & intramural leiomyomata; inactive endometrium with chronic inflammation & focal minimal; bilateral pyosalpinx & acute salpingitis Metastasis: not available	not available	not available
6G/6H	uterus adenocarcinoma	72-yr-old	Metastasis: none seen	M8140/3	182.0

	NOS and normal tissue	Caucasian			
7G/7H	uterus squamous cell carcinoma NOS and normal tissue	42-yr-old Caucasian	Metastasis: none seen	M8070/3	180.9
10G/10H	ovary serous cystadenocarcinoma NOS and normal tissue	61-yr-old Caucasian	Metastasis: metastases seen in epiploon and uterus	M8441/3	183.0
11G/11H	ovary adenocarcinoma NOS and normal tissue	41-yr-old Caucasian	Metastasis: metastases seen in adipose	M8140/3	183.0
12G/12H	ovary papillary serous carcinoma and normal tissue	72-yr-old	moderately differentiated (grade II) in left & right ovaries & fallopian tubes; no evidence of tumour in peritoneum; history of TVH (prolapse); post menopausal; on Premarin for 3 months Metastasis: metastases to liver, rectosigmoid, appendix, omentum & lymph nodes	not available	not available
14G/14H	cervix adenosquamous carcinoma and normal tissue	60-yr-old Caucasian	Metastasis: none seen	M8560/3	180.9
1J/1K	colon adenocarcinoma NOS and normal tissue	67-yr-old female Caucasian	Metastasis: none seen	M8140/3	154.1
2J/2K	colon adenocarcinoma and normal tissue	68-yr-old male	invasion seen throughout sigmoid wall to subcutaneous fat Metastasis: metastases to liver	not available	not available
3J/3K	colon adenocarcinoma NOS and normal tissue	64-yr-old male Caucasian	Metastasis: none seen	M8140/3	153.0
4J/4K	colon adenocarcinoma NOS and normal tissue	60-yr-old male Caucasian	Metastasis: none seen	M8140/3	153.9
5J/5K	colon adenocarcinoma NOS and normal tissue	61-yr-old female Caucasian	Metastasis: None seen	M8140/3	153.9
6J/6K	colon adenocarcinoma NOS and normal tissue	75-yr-old female Caucasian	Metastasis: metastases to lymph nodes	M8140/3	153.9
7J/7K	colon adenocarcinoma NOS and normal tissue	69-yr-old female Caucasian	Metastasis: none seen	M8140/3	153.9
8J/8K	colon adenocarcinoma NOS	35-yr-old female	sigmoid colon	M8140/3	153.3

	and normal tissue	Caucasian	Metastasis: metastases to lymph nodes		
9J/9K	colon adenocarcinoma NOS and normal tissue	58-yr-old male Caucasian	Metastasis: none seen	M8140/3	153.9
10J/10K	colon adenocarcinoma NOS and normal tissue	79-yr-old female Caucasian	Metastasis: none seen	M8140/3	153.9
11J/11K	colon adenocarcinoma NOS and normal tissue	51-yr-old female Caucasian	Metastasis: none seen	M8140/3	153.9
13J/13K	Lung adenocarcinoma and normal tissue	59-yr-old female Caucasian	Moderately to poorly differentiated with regions of necrosis & numerous mitoses; left upper lobe; 6 cm with focal papillary & clear cell features Metastasis: Metastases to lymph nodes	not available	not available
14J/14K	lung bronchioalveolar adenocarcinoma and normal tissue	44-yr-old female Caucasian	right lobe; bronchioalvelor pattern with high grade nuclear features (3 cm in diameter); vascular invasion seen Metastasis: metastases to lymph nodes		
15J/15K	lung squamous cell carcinoma and normal tissue	75-yr-old male	right upper lobe, posterior segment; moderately differentiated & keratinizing; angiolymphatic & perineural invasion seen; history of squamous cell carcinoma in base of tongue & left lung Metastasis: none seen	not available	not available
1M/1N	stomach cystadenocarcinoma NOS and normal tissue	35-yr-old male Caucasian	Metastasis: none seen	M8440/3	151.9
2M/2N	stomach adenocarcinoma NOS and normal tissue	62-yr-old female Caucasian	Metastasis: metastases to lymph nodes	M8140/3	151.9
3M/3N	stomach adenocarcinoma NOS and normal tissue	70-yr-old male Caucasian	Metastasis: none seen	M8140/3	151.9
4M/4N	stomach signet ring cell carcinoma and normal tissue	71-yr-old male Caucasian	Metastasis: metastases to lymph nodes	M8490/3	151.9
5M/5N	stomach adenocarcinoma NOS and normal tissue	58-yr-old female Caucasian	Metastasis: none seen	M8140/3	151.0
6M/6N	stomach adenocarcinoma NOS and normal tissue	65-yr-old male Caucasian	Metastasis: none seen	M8140/3	151.9

7M/7N	stomach lymphosarcoma NOS and normal tissue	61-yr-old male Caucasian	Metastasis: none seen	M9610/3	151.9
8M/8N	stomach adenocarcinoma NOS and normal tissue	69-yr-old male Caucasian	Metastasis: metastases to adipose of epiploon	M8140/3	151.9
10M/10N	rectum adenocarcinoma NOS and normal tissue	48-yr-old female Caucasian	Metastasis: metastases to lymph nodes	M8140/3	154.1
11M/11N	rectum adenocarcinoma NOS and normal tissue	70-yr-old male Caucasian	Metastasis: none seen	M8140/3	154.1
12M/12N	rectum adenocarcinoma NOS and normal tissue	66-yr-old male Caucasian	Metastasis: none seen	M8140/3	154.1
13M/13N	rectum adenocarcinoma NOS and normal tissue	67-yr-old female Caucasian	Metastasis: none seen	M8140/3	154.1
14M/14N	rectum adenocarcinoma NOS and normal tissue	male Caucasian	Metastasis: metastases to lymph nodes	M8140/3	154.1
15M/15N	rectum adenocarcinoma NOS and normal tissue	59-yr-old male Caucasian	Metastasis: metastases to lymph nodes	M8140/3	154.1
16M/16N	rectum adenocarcinoma NOS and normal tissue	64-yr-old male Caucasian	Metastasis: none seen	M8140/3	154.1
18M/18N	small intestine adenocarcinoma NOS and normal tissue	49-yr-old female Caucasian	Metastasis: metastases seen in lymph nodes	M8140/3	152.0

CHAPTER 3

IN SILICO STUDY OF THE NOVEL GENE CLWD

3. *In Silico* Study of the Novel Gene Clwd.

3.1 Introduction

Clwd or Close to Wasted Deletion was originally discovered upon sequencing a BAC carrying genomic sequence from the region of mouse chromosome 2 surrounding *Eef1a2* and the wasted deletion. Having using this sequence to query the Unigene EST database, a large cluster was found that correlated directly to sequence approx 10kb from the deletion. In addition, it was clear from the genomic sequence that there was a CpG island upstream of the sequence matching the EST cluster. As discussed previously, it was unclear how *Eef1a2* could be responsible for the phenotype, seen in the literature, in the immune system of the *wst/wst* mice. It had therefore become apparent that studies into the genes surrounding the wasted deletion would have to be undertaken to look for a potential second candidate gene. Previous work suggested that none of the other known genes in the region were affected in the wasted mouse[44, 46-48], and so it was a case of searching for hitherto unknown and uncharacterised genes. It was in this search that Clwd was discovered. It was known already that no predicted gene existed within the deletion other than the portion of *Eef1a2* already described and so this potential gene was the closest to the deletion other than *Eef1a2*. It thus made an excellent choice for an initial study.

This initial study took the form of a bioinformatic approach to gain as much information as possible from the various tools, databases and other resources available on the internet. In this chapter I will describe this *in silico* work as an introduction to Clwd and its potential characteristics.

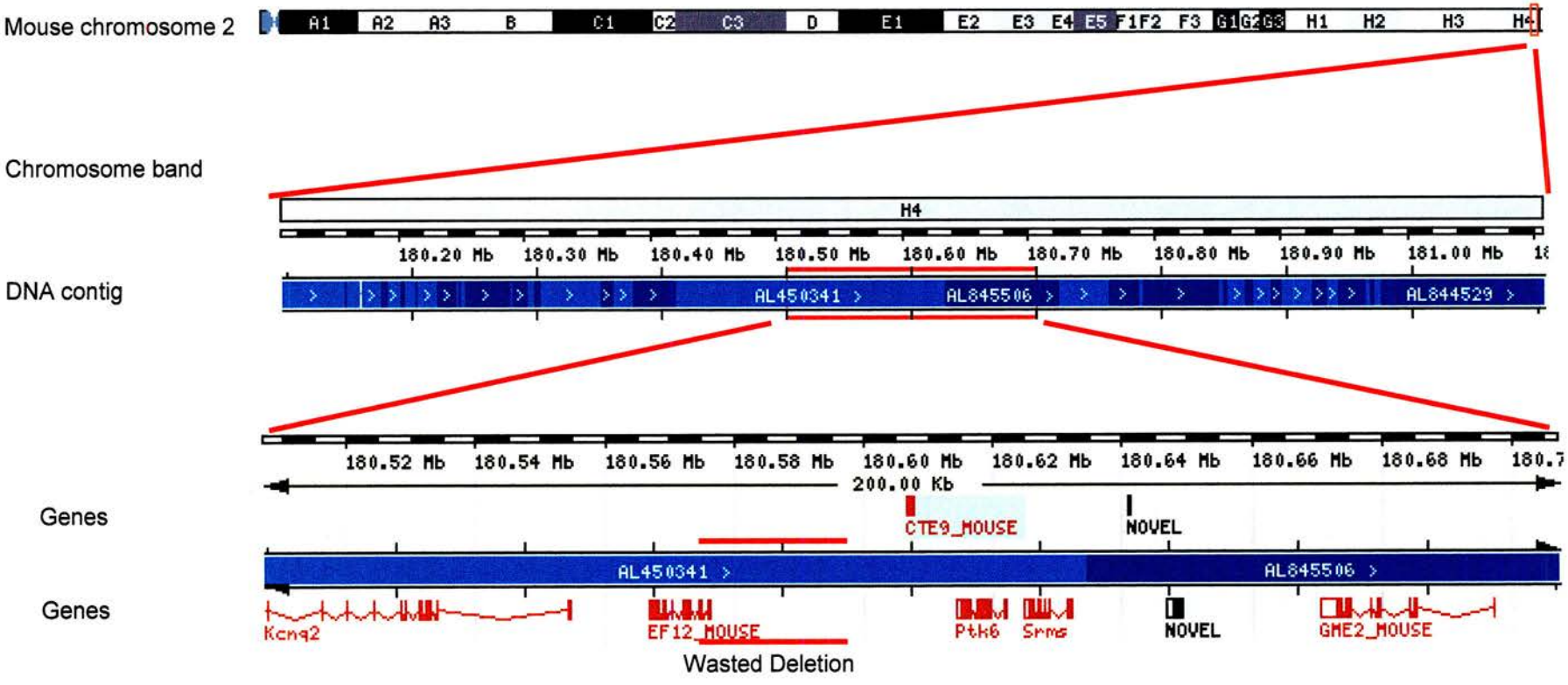


Figure 3.1 Genomic region surrounding *Eef1a2* and the wasted deletion as seen in Ensembl (www.ensembl.org). Here Clwd is seen as CTE9_MOUSE.

3.2 From EST to novel gene

3.2.1 Clwd in the EST database

The presence of ESTs or Expressed Sequence Tags can be the first hint of the presence of a novel gene as they represent the expression of the genomic DNA. On examining the genomic region in the vicinity of the wasted deletion, it became clear that there was a small region about 20kb upstream of the *Eef1a2* gene that was highly represented in databases such as Unigene at NCBI which groups ESTs that share sequence into clusters that may represent an, as yet, undiscovered novel gene. This was done before the mouse genome sequence was available, or the advent of mouse genome browsers such as that at Ensembl (www.ensembl.org) and so the identification of this candidate gene was accomplished through careful sequencing and BLAST searches of the EST data available. I shall refer to this potential gene as "Clwd" standing for "Close to the Wasted Deletion".

Within this EST cluster (now termed Mm.213943) at Unigene, it was clear that despite there being significant identity between individual sequences (enough to allow them to be confidently clustered together as representing one gene) there were some sequence differences. This took the form of short stretches *e.g.* 80-200bp that were seen in some ESTs and not in others. This suggested that if this was a *bone fide* gene, then it was one that produced alternative splice forms. On closer examination, it was clear that there were in fact five specific sequences that appeared again and again.

For the rest of this discussion I will refer to the following accession numbers as ESTs that are representative of these 5 splice forms:

Splice form	Genbank accession number
I	AK012037
II	AK014211
III	AK003282
IV	AK002602
V	AK012545

These are not isolated incidents of particular ESTs, but rather the most complete sequence representing a set of ESTs. At The Institute for Genomic Research (www.tigr.org) these sets are group into contigs with information regarding the number of ESTs in each contig.

Splice form	Representative EST	TIGR Contig	No. of ESTs
Splice form I	AK012037	TC1042785	37
Splice form II	AK014211	TC1012538	53
Splice form III	AK003282	TC1012537	40
Splice form IV	AK002602	TC1012539	42
Splice form V	AK012545	TC1012539	42

Despite the number of ESTs present in each contig it would still be very important to prove their existence as true mRNAs biologically, given the unusual splicing. This will be discussed in the next chapter.

It should also be noted at this stage that these ESTs came from a wide variety of libraries from two-cell embryo through to adult and in a wide variety of tissues. There appeared to be no strict correlation between splice form and tissue expression. Therefore it seemed that if these ESTs did represent a gene it would be one that was widely, if not ubiquitously, expressed from as early as the two-cell embryo. This would, in turn, imply a gene with a vital function necessary for life at the cellular level.

Once the genomic sequence of the mouse became available for our region of interest, it was possible to align these five ESTs with the genomic DNA sequence to see if it had the necessary characteristics to be a gene. This was done using Spidey at NCBI (National Center for Biotechnology Information) which also allowed me to assess donor and acceptor splice sites at the stretches where the sequences differed between ESTs, to see if they had the potential to be real mRNAs.

As can be seen from figures 3.2 and 3.3 below, these EST sequences are consistent with the hypothesis that they represent alternative splice forms of a genuine gene, with each isoform having legitimate splice donor and acceptor sites.

Figure 3.2a Output of NCBI's Spidey programme.
Alignment of ESTs representing Clwd splice forms with the genomic sequence from chromosome 2 (Ensembl mouse release 19.32.2). It should be noted that the sequences represented here are from the individual clones as named *i.e.* the 5' ends of the true splice forms may not be exactly as shown.

	Genomic coordinates	mRNA coordinates	Length	Donor Site	Acceptor Site
AK012037					
Exon 1	186480-186590	1-111	111	d	
Exon 2	186801-186872	112-183	72	d	a
Exon 3	186945-187022	184-261	78	d	a
Exon 4	187140-187236	262-358	97	d	a
Exon 5	187318-187614	359-655	297		a
AK014211					
Exon 1	186482-186590	1-109	109	d	
Exon 2	186801-187022	110-331	222	d	a
Exon 3	187140-187236	332-428	97	d	a
Exon 4	187318-187642	429-753	325		a
AK003282					
Exon 1	186478-186590	1-113	113	d	
Exon 2	186686-186872	114-300	187	d	a
Exon 3	186945-187022	301-378	78	d	a
Exon 4	187140-187236	379-475	97	d	a
Exon 5	187318-187640	476-798	323		a
AK002602					
Exon 1	186431-186590	1-160	160	d	
Exon 2	186686-187022	161-497	337	d	a
Exon 3	187140-187236	498-594	97	d	a
Exon 4	187318-187642	595-918	324		a
AK012545					
Exon 1	186671-187022	1-352	352	d	
Exon 2	187140-187236	353-449	97	d	a
Exon 3	187318-187640	450-772	323		a

Figure 3.2b Output of NCBI's Spidey programme

Table showing relationship of EST sequences to genomic DNA including splicing information.

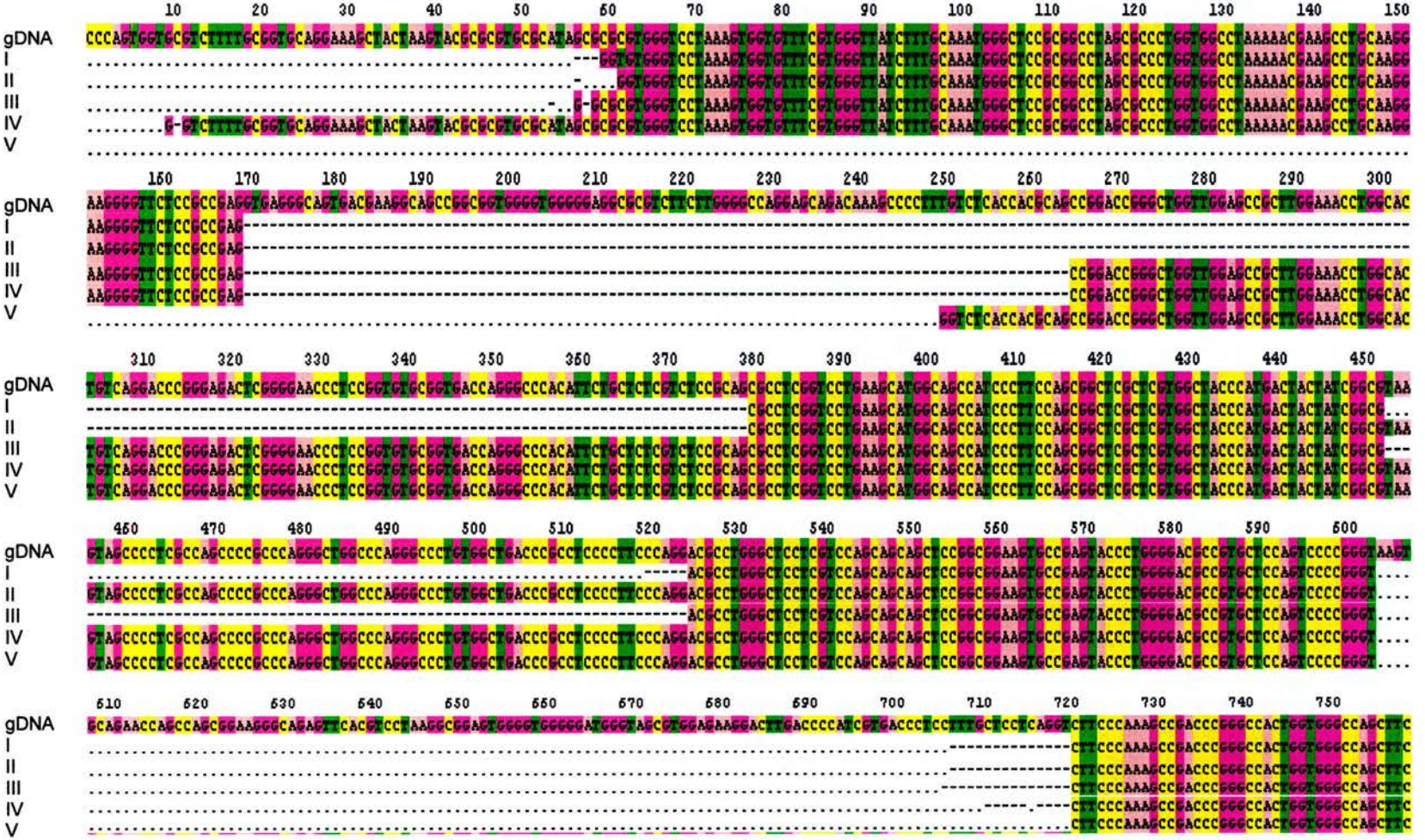


Figure 3.3 Multiple alignment of 5 splice forms and genomic sequence (continued overleaf)

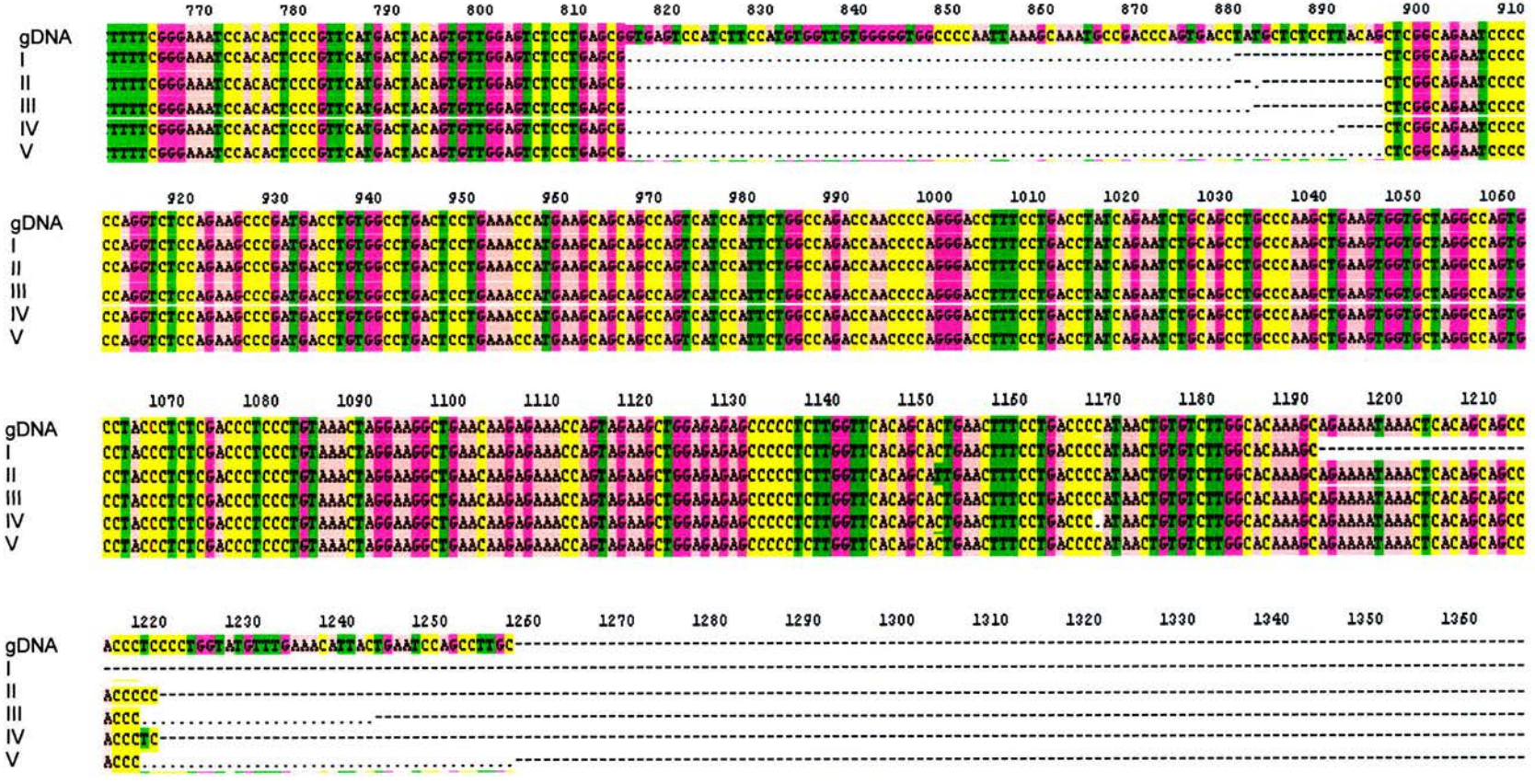


Figure 3.3 Multiple alignment of 5 splice forms and genomic sequence

3.2.2 Clwd according to gene prediction programmes

With evidence that this region did harbour a gene I ran some gene-prediction programmes on the genomic sequence in this region. One of the most valuable of these was not just a gene-prediction programme, but in fact a means for running a suite of packages over the same genomic region run at HGMP called NIX (Nucleotide Identify X). These included programmes for predicting/displaying: CpG islands, promoter regions, exons, genes, polyadenylation signal as well as BLAST-hits to trEMBL, Swissprot and Unigene. NIX, however, does not just run one, but several programmes available for each of the above functions allowing for increased confidence of a correct prediction where more than one programme picks up the same feature. See figure 3.4.

The output for Clwd confirmed the intron-exon structure I had surmised from comparing the EST database to the genomic sequence available, which can be seen by looking at the output for GRAIL, FGenes and GENSCAN in particular. These programmes compare known ESTs to the genomic sequence. They did not recognise the alternative splicing that I had noticed because these programmes only "see" alternative splice forms represented by ESTs when the splicing takes the form of skipped exons *e.g.* in EST1 exon A, B and C are present while EST2 has only exon A and C. This would be called as an alternatively spliced exon. However the alternative splice forms seen in the case of Clwd make use of alternative 5'and 3' donor sites in the first two exons, and exhibit the retention of the second intron. This type of splicing is not recognised by these programmes as it is more likely to be due to a mistake in the alignment or in the ESTs. Although I had confirmed there was more than one of each type of splice form represented in the EST database, this did emphasise again the importance of following up this work with biological experiments.

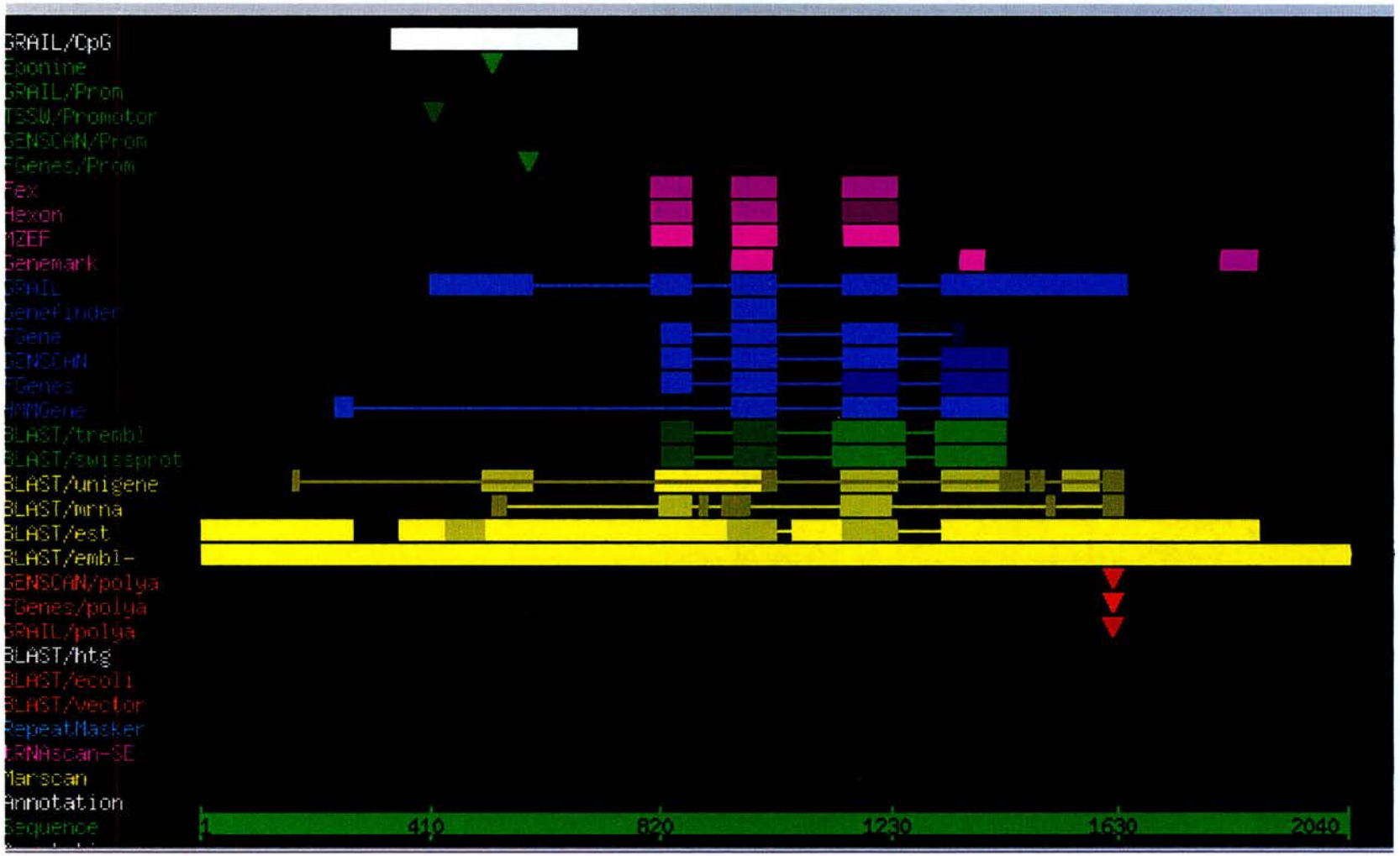


Figure 3.4 NIX output for area surrounding Clwd.

To facilitate the process of identifying a potential new gene, NIX also provides programmes to ascertain the presence of CpG islands, promoter sequences and poly A tail leader sequences whose presence would make any gene prediction from EST-genomic sequence alignment all the more sound.

For Clwd, NIX predicted these features as summarised below:

Feature	Programme	Quality of prediction
CpG Island	(GrailEXP) CpG island	Excellent
Promoter	Eponine	Good
	GRAIL/Promoter	None
	TSSW/Promoter	Marginal
	GENSCAN/Promoter	None
	FGenes/Promoter	Good
Poly-A Tail	GENSCAN/polya	Good
	FGenes/polya	Excellent
	GRAIL/polya	Good

While three of the five promoter prediction programmes found a promoter upstream of the starting codon of Clwd, it is a little disconcerting that two programmes did not. However, as the two programmes that did not find a promoter are usually used for gene prediction only, I decided to try out some other internet-accessible promoter-predictors. First I tried the Dragon Promoter Finder at The Institute for Infocomm Research, Singapore (http://research.i2r.a-star.edu.sg/promoter/promoter1_4/DPF.htm). This predicted a promoter sequence upstream of the start codon with 65% sensitivity. The Neural Network Promoter Predictor (www.fruitfly.org/seq_tools/promoter.html) gave a positive result also with a score of 0.99. This score is related to the confidence of the prediction as shown.

Threshold	% promoters recognised	False positives
0.99	10%	0.0%
0.97	20%	0.0 – 0.1%
0.92	30%	0.1 - 0.3%
0.85	40%	0.1 – 0.4%
0.7	50%	0.8 – 1.0%

Having used these additional promoter-predictions I was more confident that there was a real promoter to allow for the transcription of the gene.

This combination of confirming the intron-exon structure and predicting various features of a gene model such as CpG island, promoter and polyadenylation sites suggested that Clwd was a previously undescribed gene.

Another feature of Clwd I wanted to look at in a genomic context was to see if there were any pseudogenes of Clwd that I could detect, and whether these were expressed in any way, contributing to the EST data. Using the genomic BLAST service at Ensembl I found that there were three areas of the genome with significant similarity to the Clwd region of chromosome 2. These were located on the X chromosome, chromosome 14 and chromosome 10. An overview of their location and an alignment of these sequences are shown below.

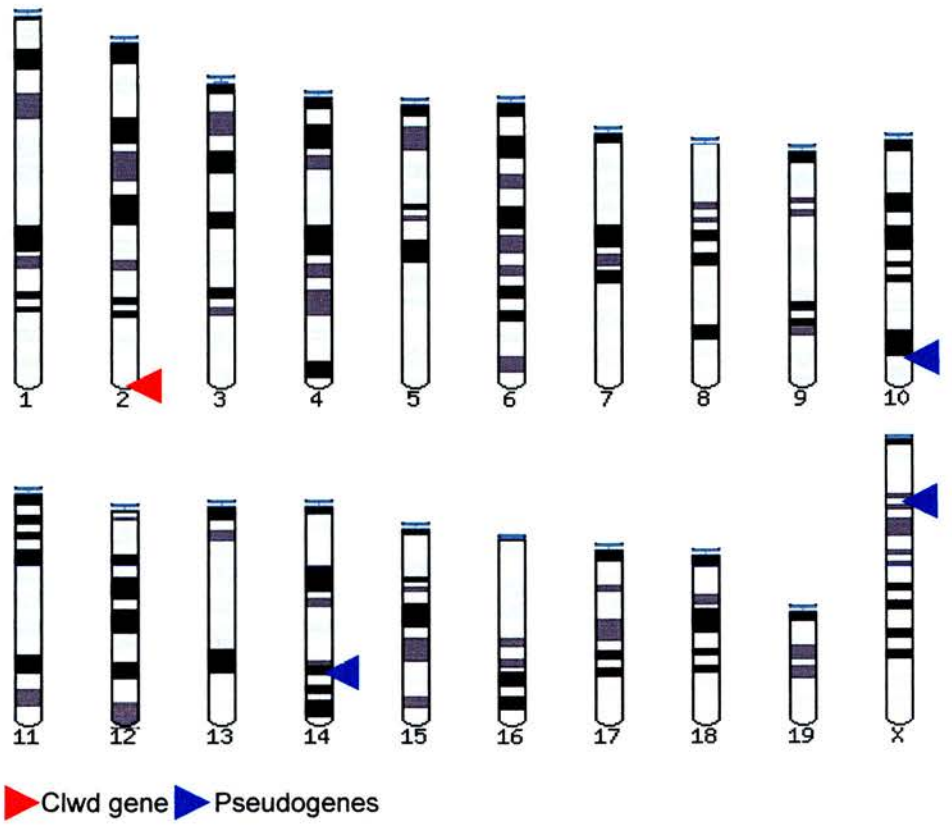


Figure 3.5 Illustration of the position of the pseudogenes of Clwd in the mouse genome

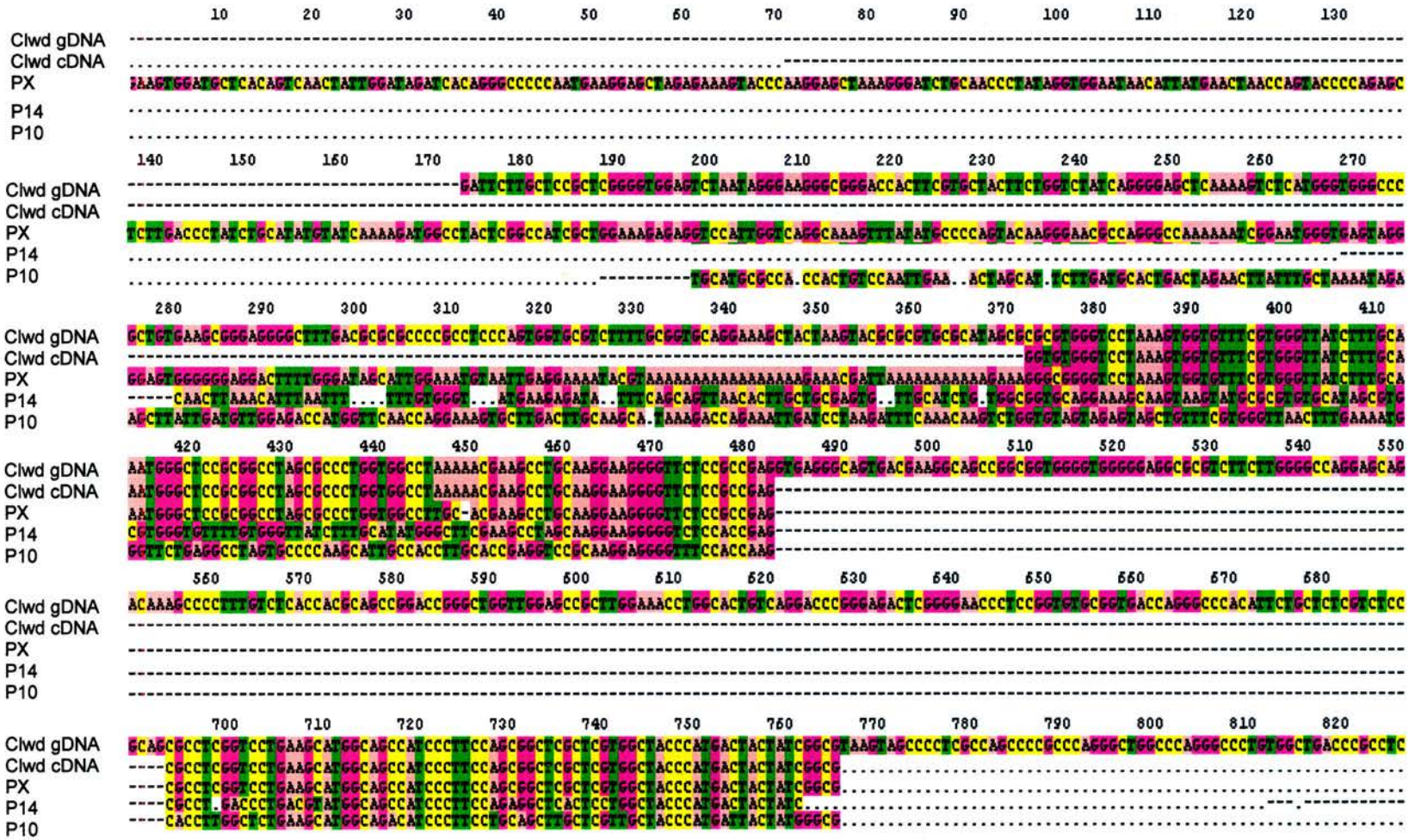


Figure 3.6 Clustal W alignment of Clwd genomic DNA sequence (Clwd gDNA), Clwd splice variant AK012037 (Clwd cDNA), and the pseudogenes of Clwd on chromosome X (PX), 14 (P14) and 10 (P10). (Contd. overleaf)

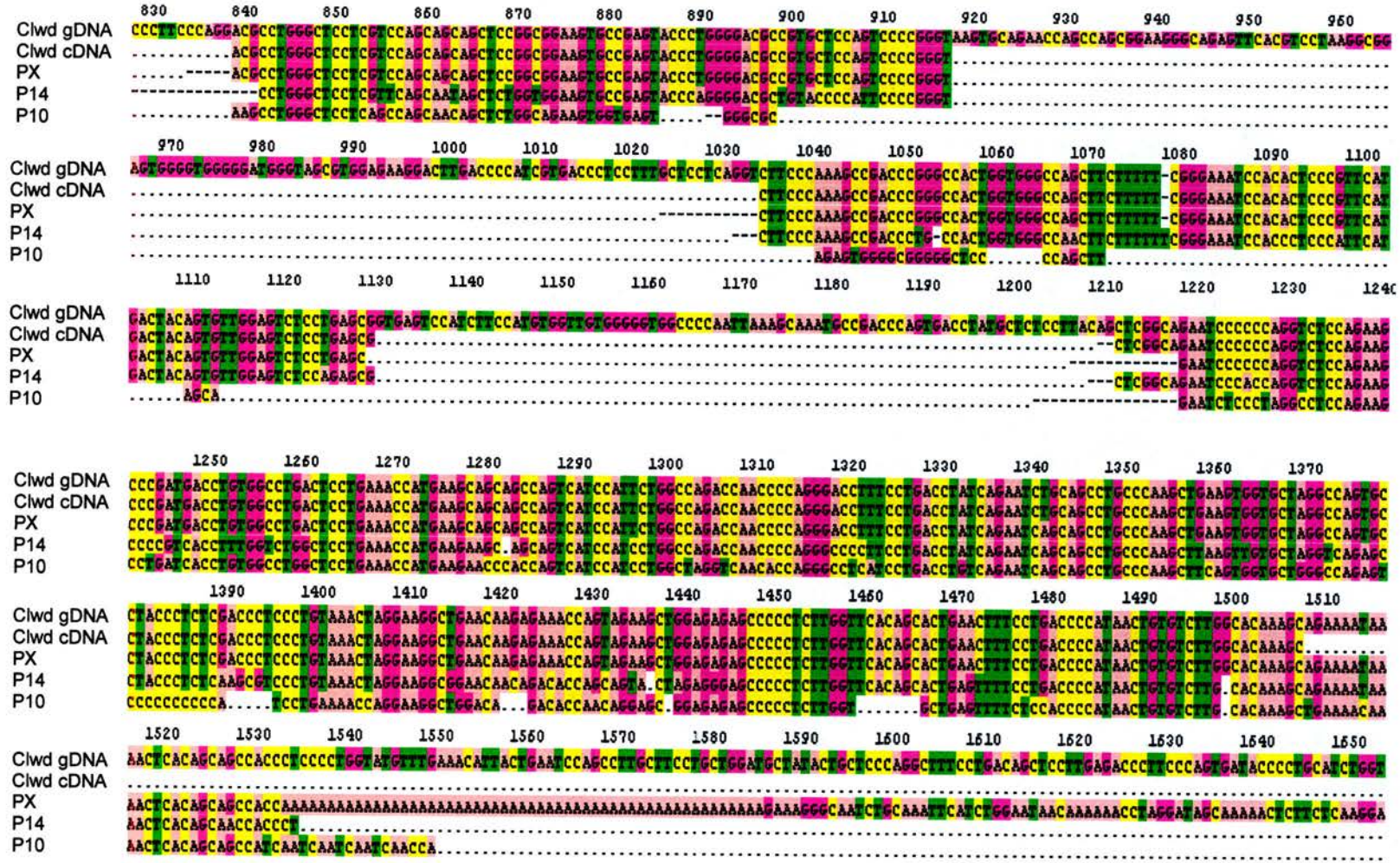


Figure 3.6 Clustal W alignment of Clwd genomic DNA sequence (Clwd gDNA), Clwd splice variant AK012037 (Clwd cDNA), and the pseudogenes of Clwd on chromosome X (PX), 14 (P14) and 10(P10)

This alignment shows that the three pseudogenes shown are indeed processed versions of Clwd lacking in introns. However, there are also deletions and mutations in these versions of Clwd that means that none produce the same open reading frame as Clwd. In addition, the open reading frame of each, shown below, when used as a query of the mouse EST database in a tBLASTn search fails to find any EST with the same sequence, implying that these pseudogenes are silent.

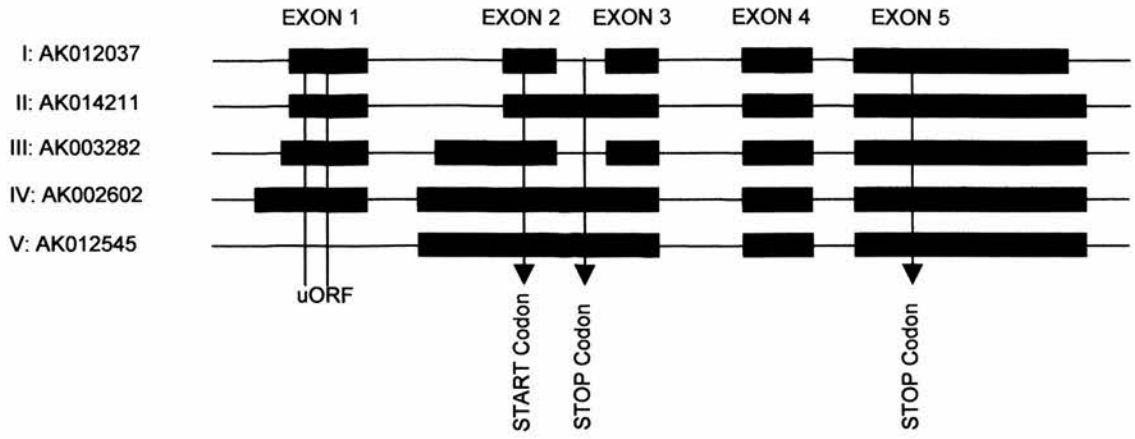
```
Clwd
MAAIPSSGSLVATHDYYRRRLGSSSSSSSGGSAEYPGDAVLQSPGLPKADPGHWWAS
FFFGKSTLPFMTTVLESPERSAESPQVSRSPMTCGLTPETMKQQPVIHSGQTNPRDL
S*

Chromosome X translated
MAAIPSSGSLVATHDYYRRRLGSSSSSSSGGSAEYPGDAVLQSPGLPKADPGHWWAS
FFFGKSTLPFMTTVLESPERIPPGLQKPDDLWPDS*

Chromosome 14 translated
MAAIPSRGSLLATHDYYPWAPRSAIALVEVPSTQGTLYPIPRVFPKPTLPLVGQLLF
SGNPPSHS*

Chromosome 10 translated
MADIPSCSLLVATHDYYGRSLGSSASNSSGRSGEWAQSGAGAPQLSRISLGLQKPDH
LWPGS*
```

Figure 3.7 ORFs of Clwd and three pseudogenes on chromosome X, 14 and 10

3.2.3 Open reading frames of Clwd ESTs

The next important stage of data mining the genomic sequence of Clwd was to look at the potential protein produced by this gene. The open reading frame of Clwd can be seen in figure 3.8. This figure also shows that the splice variants of Clwd produce one of two types of ORF depending on whether or not the second intron is retained. The other differences between the splice variants occur in the first exon and intron that, in the case of Clwd lie in the 5' untranslated region.
The two types of ORF are essentially a full-length and truncated version of the same open reading frame as a stop codon is present within the second intron

In addition to Clwd's main ORF, it also exhibits a short conserved upstream ORF as shown in figure 3.9. This feature would not have been noticed had it not been conserved through evolution. Upstream open reading frames are just beginning to be understood. They appear to be important in the post-transcriptional control of gene expression in that their presence can repress the translation of the main open reading frame.



Splice forms I and III would produce

MAAIPSSGSLVATHDYYRRRLGSSSSSSSGGSAEYPGDAVLQSPGLPKADPGHWWAS

FFFGKSTLPFMTTVLESPERSAESPQVSRSPMTCGLTPETMKQQPVIHSGQTNPRDL

S*

Splice forms II, IV and V would produce

MAAIPSSGSLVATHDYYRRK*

Isoform I: AK012037 – Full Length CLWD

```
GGTGTGGGTC CTAAAGTGGT GTTTCGTGGG TTATCTTTGC AA ATG GGC TCC GCG GCC TAG
                                                  M   G   S   A   A   *
CGCCCTGGTG GCCTAAAAAC GAAGCCTGCA AGGAAGGGGT TCTCCGCCGA GCGCCTCGGT
CCTGAAGC                                        ATG GCA GCC ATC CCT
                                                 M   A   A   I   P
TCC AGC GGC TCG CTC GTG GCT ACC CAT GAC TAC TAT CGG CGA CGC CTG GGC TCC
 S   S   G   S   L   V   A   T   H   D   Y   Y   R   R   R   L   G   S
TCG TCC AGC AGC AGC TCC GGC GGA AGT GCC GAG TAC CCT GGG GAC GCC GTG CTC
 S   S   S   S   S   G   G   S   A   E   Y   P   G   D   A   V   L   Q
CAG TCC CCG GGT CTT CCC AAA GCC GAC CCG GGC CAC TGG TGG GCC AGC TTC TTT
 S   S   P   G   L   P   K   A   D   P   G   H   W   W   A   S   F   F
TTC GGG AAA TCC ACA CTC CCG TTC ATG ACT ACA GTG TTG GAG TCT CCT GAG CGC
 F   G   K   S   T   L   P   F   M   E   S   P   E   R   S   A   E   S
TCG GCA GAA TCC CCC CAG GTC TCC AGA AGC CCG ATG ACC TGT GGC CTG ACT CCT
 P   Q   V   S   R   S   T   T   V   L   P   M   T   C   G   L   T   P
GAA ACC ATG AAG CAG CAG CCA GTC ATC CAT TCT GGC CAG ACC AAC CCC AGG GAC
 E   T   M   K   Q   Q   P   V   I   H   S   G   Q   T   N   P   R   D
CTT TCC TGA
 L   S   *
CCTATCAGAA TCTGCAGCCT GCCCAAGCTG AAGTGGTGCT AGGCCAGTGC CTACCCTCTC GACCCTCCCT
GTAAACTAGG AAGGCTGAAC AAGAGAAACC AGTAGAAGCT GGAGAGAGCC CCCTCTTGGT TCACAGCACT
GAACTTTCCT GACCCCATAA CTGTGTCTTG GCACAAAGC
```

Figure 3.8 The open reading frames produced by Clwd splice forms

3.2.4 Conservation of Clwd through evolution

Another important piece of evidence in favour of Clwd being a true gene is the fact that it is conserved from fish to human. We assume that DNA sequence that is conserved through several species is important for normal function as otherwise, over time, it would have accumulated mutations and/or become rearranged. Therefore the fact that Clwd is present through several species is a sign of its having a function important to those species. Clwd cannot be detected in either *D. melanogaster* or *C. elegans*, thus it seems to be important to vertebrates only.
It can also be seen that the upstream open reading frame present in mouse is also present in each of these species. While the sequence is well conserved in zebrafish, xenopus, rat, mouse and human, the fugu sequence is quite different except for the last position, which, while not the alanine of the other species, is isoleucine, another hydrophobic residue.

(A) Upstream open reading frame

Colour	Characteristic	Amino Acid
Blue	Hydrophobic	ACFHILMVWY
Magenta	Negatively charged	DE
Red	Positively charged	KR
Green	Polar	STQN
Pink		C
Orange		G
Yellow		P
Cyan	Aromatic	FYW

(B) Main open reading frame

Figure 3.9 Multiple alignment of the upstream and main open reading frames of Clwd showing conservation of amino acid sequence through evolution

3.3 Identification of a potential role for Clwd

3.3.1 Potential characteristics of Clwd

Having determined the gene structure, splice variants and peptide sequence of Clwd, the next step was to see what biological information could be gleaned from this raw data.

With the peptide sequence, I used bioinformatic tools to get information on the subcellular localisation, post-translational modification and/or conserved domains that could give a clue as to the function of Clwd

First, I tried to localise Clwd to a particular cell organelle/compartment by using the PSORT programme at HGMP. This assesses Clwd for the presence of various signal sequences that are known to direct the location of known proteins. In order to get an impression of the value of this prediction I used the peptide sequences of Clwd from different species in this programme.

Cell compartment	Human (%)	Mouse (%)	Rat (%)	Xenopus (%)	Zebrafish (%)	Fugu (%)
Nuclear	28	32	40	28	20	28
Cytoplasmic	16	32	24	24	24	24
Mitochondrial	40	20	20	36	44	36
Cytoskeletal	4	12	8	8	4	4
Vacuolar	-	4	-	-	-	-
Peroxisomal	4	-	-	-	4	4
Plasma membrane	4	-	-	-	-	-
ER	4	-	-	-	-	-
Secretory vesicular	-	-	8	4	4	-
Extracellular	-	-	4	-	-	4

Although there is some agreement between the various species' Clwd subcellular localisation, the predictions are far from definite, with nuclear, cytoplasmic and mitochondrial all being quite close for each species.

However another programme, Target P (http://www.cbs.dtu.dk/services/) gave the following result:

Name	mTP	SP	Other	Location	RC
Human	0.780	0.072	0.177	Mitochondria	2
Mouse	0.802	0.054	0.200	Mitochondria	2
Rat	0.755	0.053	0.217	Mitochondria	3
Xenopus	0.745	0.029	0.307	Mitochondria	3
Zebrafish	0.715	0.024	0.399	Mitochondria	4
Fugu	0.814	0.022	0.285	Mitochondria	3

mTP - mitochondrial targeting peptide
SP - signal peptide
RC – reliability class - measure of the size of the difference (diff) between the highest (winning) and the second highest output scores
RC 1: diff > 0.800, RC 2: 0.800 > diff > 0.600, RC 3: 0.600 > diff > 0.400, RC 4: 0.400 > diff > 0.200, RC 5: 0.200 > diff

This suggests that Clwd is in fact mitochondrial for all species.
On consulting a third programme however, MitoProt II 1.0, which uses the sequence of the protein to predict the probability of that protein being exported to the mitochondria the results were as follows.
Probability of export to mitochondria:

Mouse	0.2217
Human	0.2093
Rat	0.1876
Xenopus	0.5416
Fugu	0.5155
Zebrafish	0.3376

It can be easily seen from these results, therefore, that it is difficult to conclude where exactly Clwd is located given that its sequence doesn't seem to have any of the definitive patterns or signals that can be interpreted easily by computer

programmes. All that can be said is that there is a chance that Clwd is expressed in the mitochondria.

In addition to determining subcellular localisation, there are also many programmes available that predict the presence of various forms of post-translational modification. The post-translational modification of a protein has an important effect upon characteristics such as subcellular localisation, function and interaction with other proteins.

Phosphate, glycosyl or myristyl groups modify particular amino acids within a peptide sequence. It is the context within which these amino acids find themselves that dictates if the modification will occur. Programmes that predict these modifications use this to calculate the probability that the amino acid is targeted.

One of the first programmes I ran on Clwd combined a signal peptide identifier, which should predict if Clwd is exported out of the cell and thus more likely to have modifications like glycosylation, and a glycosylation prediction programme. This was the NetOGlyc 3.0 Server (http://www.cbs.dtu.dk/services/). This gave the following results for mouse Clwd.

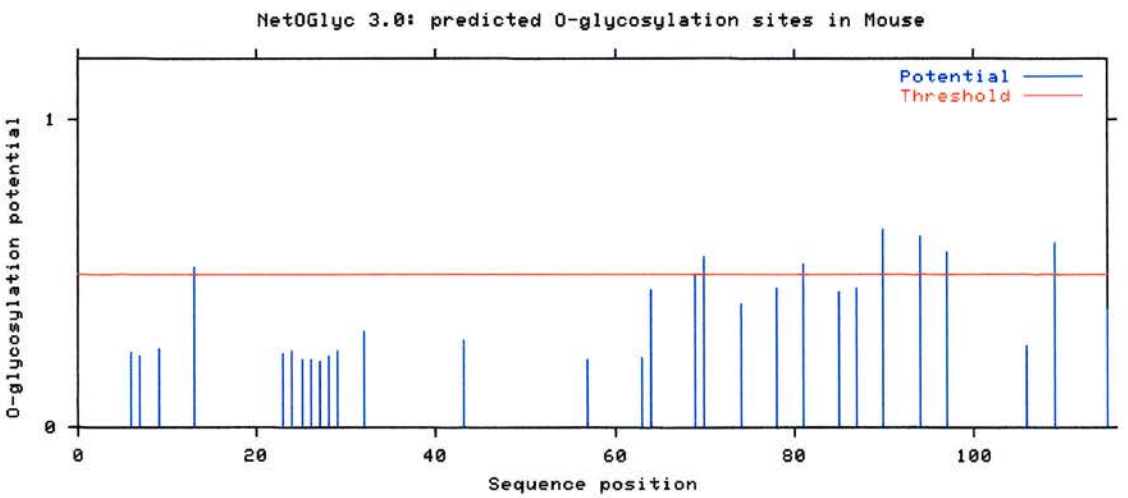


Figure 3.10 Graphic output from NetOGlyc 3.0 programme predicting presence of O- glycosylation sites in Clwd

This shows the potential for O-glycosylation of each serine/threonine residue in the sequence. Those lines that cross the threshold are more likely to be modified *in vivo*, although in this case, these lines are not much higher than the threshold. This programme also looked for a signal peptide that it did not find in Clwd and so the following accompanied the above result:

```
WARNING: This sequence may not contain a signal peptide!!

Proteins without signal peptides are unlikely to be exposed to
the O-glycosylation machinery and thus may not be glycosylated
(in vivo) even though they contain potential motifs.

SignalP-NN euk predictions are as follows:
```

Cmax	Position	?	Ymax	Position	?	Smax	Position	?	Smean	?
0.131	34	No	0.055	16	No	0.318	1	No	0.143	No

C-score	raw cleavage site score The output score from networks trained to recognise cleavage sites vs. other sequence positions.
S-score	signal peptide score The output score from networks trained to recognise signal peptide vs. non-signal-peptide positions.
Y-score	combines the above to find the total cleavage site score

This suggests that Clwd is not O-glycosylated or, for the same reason, N-glycosylated.

There is a form of glycosylation that occurs in nuclear and cytoplasmic proteins - this is the modification of serine/threonine residues by O-linked N-acetylglucosamine. This is a dynamic modification, more similar to phosphorylation than traditional O-glycosylation. Considering Clwd is unlikely to have a signal peptide, this form of modification is far more likely than traditional glycosylation, and indeed when I used YinOYang 1.2 (http://www.cbs.dtu.dk/services/) the results confirmed this likelihood.

In addition to showing potential O-GlcNAc sites in a peptide, this programme also runs "NetPhos" which predicts which serine/threonine/tyrosine residues will be phosphorylated. As a result the residues that have the potential to be both phosphorylated and O-GlcNAcylated are displayed. This is important as there are proteins known to have both modifications at one site in their sequence and that these are reversibly and dynamically modified at different times in the cell.

The raw output for Clwd is shown in the next figure. It can be easily seen that Clwd could well be both phosphorylated and O-GlcNAcylated *in vivo*. In addition, on comparing the sites modified on each species' Clwd, the first two regions of modification occur in each, as well as a broad correlation in modification along the length of the protein.

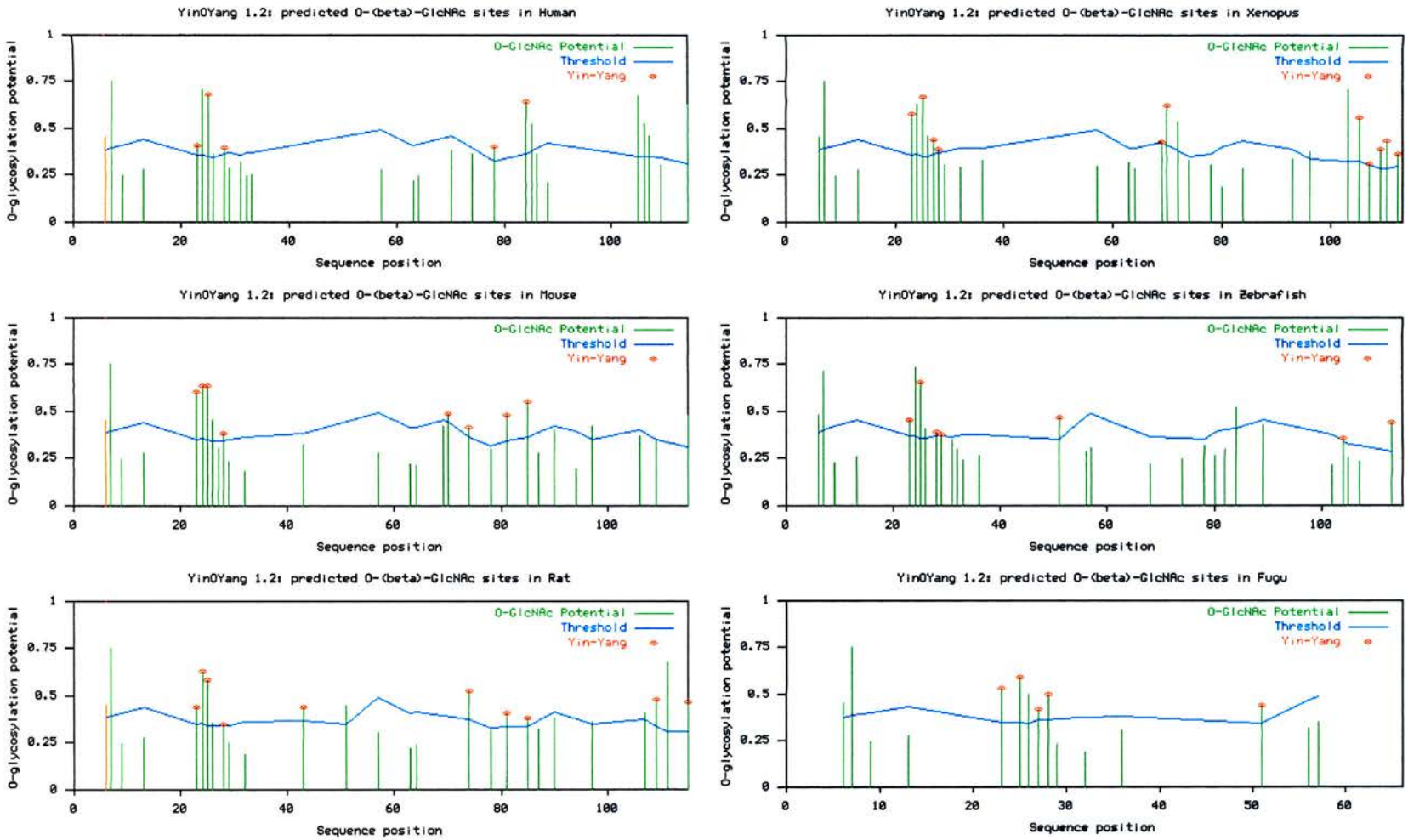


Figure 3.11 Graphical output of YinOYang 1.2 programme showing the presence of O-GlcNAcylation and phosphorylation sites in known Clwd sequences

G= O-GlcNAc
Y= YinYang *i.e.* O-GlcNAc and Phos.
O-GlcNAc predicted: *(different strengths)*
G/Y Potential > Thresh-1
G/Y Potential > Thresh-2 (Thresh-2 is a threshold based on more stringent surface measures)
G/Y Potential > (Thresh-2 + 0.1)
G/Y Potential > (Thresh-2 + 0.1) AND Potential >= 0.75

Figure 3.12 Illustration of correlation of O-GlcNAcylation and phosphorylation sites in known Clwd sequences

Other post-translational modifications that could affect Clwd are GPI (glycosylated phosphatidylinositol) anchors that attach proteins to the cell membrane; N-myristoylation sites direct and anchor proteins to membranes and tyrosine sulphation that occurs on proteins that go through the secretory pathway. Clwd is not predicted to have any of these modifications using the primary sequence alone.

Using several other programmes (PredictProtein http://cubic.bioc.columbia.edu/predictprotein/ and REP http://www.embl-heidelberg.de/~andrade/papers/rep/search.html at EMBL) I discovered that Clwd is unlikely to have transmembrane helices or repeats. The predicted pI and Mw for the various forms of Clwd are as shown.

Species	pI	Mw (kD)
Human	6.82	12.26
Mouse	6.89	11.78
Rat	7.97	12.09
Xenopus	7.85	11.96
Fugu	8	7.21
Zebrafish	7.84	12.46

In addition to motifs for post-translational modification, the primary sequence of a peptide can also have particular patterns or motifs that can designate some function or potential interaction of the peptide. These include such things as DNA binding domains, ATP binding domains, *etc.* I used the Prosite database of protein families and domains at Expasy (www.expasy.ch) to scan Clwd for any domains that may be present.

The patterns that were found were: N-glycosylation sites, Casein kinase II phosphorylation sites, Protein kinase C phosphorylation sites, N-myristoylation sites and ATP/GTP-binding site motif A (P-loop) with the following distribution.

Species	N-glycosylation	Casein kinase II phosphorylation	Protein kinase C phosphorylation	N-myristoylation	ATP/GTP binding
Human	Yes	Yes	No	Yes	Yes
Mouse	No	Yes	Yes	Yes	Yes
Rat	Yes	No	Yes	Yes	Yes
Xenopus	No	Yes	Yes	Yes	Yes
Zebrafish	Yes	Yes	Yes	Yes	No
Fugu	No	No	No	Yes	No

```
Human      MAAIPSSGSLVATHDYYRRRLGSTSSNSSCSSTECPGEAIPHPPGLPKADPGHWWASFFFGKSTLPFMATVLE
Mouse      MAAIPSSGSLVATHDYYRRRLGSSSSSSGGSAEYPGDAVLQSPGLPKADPGHWWASFFFGKSTLPFMTTVLE
Rat        MAAIPSSGSLVATHDYYRRRLGSSSSNSSGGSAEYPGDAVPHSPGLPKADSGHWWASFFFGKSTLPFMAAVLE
Xenopus    MAAIPSSGSLVATHDYYRRRLGSTSSSSSCGSVDYSGEVIPHHPGLPKADPGHWWASFFFGKSTHPVMTTVSE
Zebrafish  MAAIPSSGSLIATHDYYRRRIGSTSSNSSCSSSEYTGEVIPHPPGLARQDSGHWWSSFFFGKQNQMGTPNGFE
Fugu       MAAIPSSGSLVATHDYYRRRLGSASSSSSCGSAEYTGEVIPHHPGLPRQDSGHWWTSFFFAKQNQP

Human      SAEHSEPPQASSSMTACGLARDAPRKQPGGQSSTASAGPPS
Mouse      SPERSAESPQVSRSPMTCGLTPETMKQQPVIHSGQTNPRDLS
Rat        SPEHSAESPQASRSPISCGLAPETMKKQPVMHPSQTNSRAPS
Xenopus    SPENSGSFRITNGLVPCGLTQESVQKQKVSDSKSNSSPSA
Zebrafish  SQQKTGTYTVTNGQVTCVAREIVMKRHLSESSDSGKEPGSPL
Fugu
```

N-glycosylation
Casein kinase II phosphorylation
Protein kinase C phosphorylation
N-myristoylation
ATP/GTP-binding site

However each of these patterns comes with the following note [Warning: pattern with a high probability of occurrence]. This means the consensus sequence is short and so the probability of it occurring in any sequence by chance is quite high. For example we know already that Clwd does not have a signal peptide and so is unlikely to be N-glycosylated and yet a motif that is glycosylated is simply "– NSSC–" a sequence quite likely to occur by chance.

Also, because none of the myristoylation recognition sites are found at the N-terminus of the protein we suspect that Clwd would not be modified by a myristyl group at an internal recognition site.

One of the more interesting patterns here is the possibility that Clwd can bind ATP/GTP. Up until the last version of Prosite, this pattern was not deemed as having a high probability of occurrence. As of Dec 2003 however, this has changed and so one must be wary of referring to Clwd as an ATP/GTP binding protein without biological evidence. It is also worth noting that despite the sequence being well conserved in that area between the fishes and higher vertebrates, there are enough

changes for this area to no longer contain an ATP-binding site according to Prosite *i.e.* "-ASFFFGKS-" is a motif but "-SSFFFGKQ-" is not. It could be possible that Clwd has acquired a different function in the higher vertebrates as sequence changes result in the appearance of new functional motifs, or this ATP/GTP-binding site could be a spurious result.

3.3.2 Relationship to known proteins

Having used the programmes available to get some basic information about Clwd, the next step was to use BLAST (Basic Local Alignment Search Tool) available at NCBI, HGMP and Expasy to look for similar amino acid sequences held in the Swiss-prot and trEMBL databases. I hoped this would show me if Clwd was similar in any way to known proteins other than in the context of well-characterised domains.
However when I use BLAST to search the protein databases, the only significant hits are to Clwd itself in each of the species already described.
I also used PSI-BLAST that uses an iterative approach to identify more distantly related proteins. Unfortunately this did not result in any more significant hits.

I therefore used a BLAST search available at NCBI that looks for short nearly exact matches. Here, in addition to finding Clwd from other species I also came across other hits. It is difficult to interpret these results. As they are short, good matches, an example of which is shown below, it could represent a pattern or domain that has yet to be described. However it should also be noted that most of these matches occurred in regions where the amino acid composition was somewhat skewed toward one particular residue, and so what the BLAST search has really done is found serine, or proline rich regions see Results [1]. There were some non-biased matches though see Results [2]
As can be imagined considering these could be new domains, the proteins that were hit are not well described in the literature or databases and the subject sequence does not represent any known domain.

Thus it seems that not only is Clwd a novel gene, but in fact a pioneer gene – unlike anything already known.

Results [1].

Microphthalmia-associated transcription factor:

Similar to spermatid-specific thioredoxin [Rat]

Results [2]

G protein-coupled receptor LPSenhR-1 [*Oncorhynchus mykiss*]

Figure 3.13 Results of using BLAST in "short nearly exact matches" format to look for Clwd homology (multiple alignment in Clustalx colour).

3.3.3 Where is Clwd important?

One of the newer tools available on the internet is one that can query the SAGE data being generated at the NIH. SAGE or Serial Analysis of Gene Expression is a technique that produces a quantitative profile of cellular gene expression. For an in-depth description of the method used in SAGE see chapter 2.A

Essentially, the SAGE technique quantifies a "tag" which represents a gene transcript. A tag, for the purposes of SAGE, is a nucleotide sequence of a defined length, directly 3'-adjacent to the 3'-most restriction site for a particular restriction enzyme. By being able to find the tag representing a gene of interest and using that to query the database, one can obtain a "virtual Northern blot" of the gene. As this database was originally set up as part of the Cancer Genome Anatomy Project it is also possible to compare the expression of a human tag/gene in normal and cancerous tissues and/or cell lines.
Both human and mouse Clwd are represented by tags in this database and so I could obtain a description of Clwd gene expression using SAGE, see figure 3.14a, b and c. This again shows the wide expression suggested by the EST database with Clwd seen in the mouse brain, heart, liver, testis, spleen, limbs and skin; as well as ES cells, embryonic carcinoma cells, T cells and dendritic cells. Expression seems particularly high in the fore and hind limbs, a CD4+ T cell line and liver.

As for the human cancer genome information, there appears to be an increase in Clwd expression in tumours of the breast, prostate, stomach, pancreas, peritoneum and glioblastoma. It is worth noting that Clwd lies on human chromosome 20 at 20q13.3. 20q12-13 amplifications are found frequently in ovarian tumours associated with poorer survival and a more aggressive tumour pathology and in a variety of other solid tumours, most notably breast and colorectal cancer.

Library name	Tags per million	Tag counts	Total tags
mouse forelimb	263	18	68405
mouse hindlimb	204	14	68450
Cardiac fibroblast	90	10	110169
SAGE Mouse GCP control granular cell normal cerebellum cell line neonate EST	102	9	87837
P19 embryonic carcinoma (EC) cells, day 3+3.0	169	9	53022
Treg line SKA from skin graft	362	8	22096
P19 embryonic carcinoma (EC) cells, day 3+0.5	156	8	50983
SAGE mouse GCP+SHH granular cell normal cerebellum cell line neonate EST	81	7	85510
R1 Embryonic Stem Cells	50	7	137906
mouse liver: C57BL/6Jlco (B6)	266	5	18743
mouse liver: apoE3L transgenics	217	4	18419
ES derived DCs (LPS treated)	126	4	31681
SAGE Medullo 3871 brain medulloblastoma CGAP non-normalized SAGE library method bulk	92	4	43476
SAGE MouseP8 PGCP brain normal bulk CGAP non-normalized SAGE library method bulk	48	3	61562
Undifferentiated P19 embryonic carcinoma (EC) cells	46	3	64952
bmDC (mouse)	137	3	21781
bmDC (IL 10 treated)	98	3	30407
ES derived DCs	65	2	30696
Adult testis somatic cells	24	2	81977
fetal testis somatic cells	30	2	65593
CD4+CD25neg spleen	85	2	23343
CD4+CD25neg activated	93	2	21293
Treg line SKB from skin graft	99	2	20128
Tr1D1 (CD3 activated)	100	2	19975
Adult mouse heart	23	2	84192
ESF 116	66	1	15094
SAGE P23 cerebellum	51	1	19436
Treg clone Tr1D1 pure	82	1	12101
bmDC (IL 10+LPS)	32	1	31104
bmDC (Vitamin D3 treated)	48	1	20452
mouse hippocampus: short attack latency (SAL) mice	32	1	31205
bmDC (LPS treated)	76	1	13073

Figure 3.14a Output from SAGE at NCBI for tag representing mouse Clwd - Virtual Northern

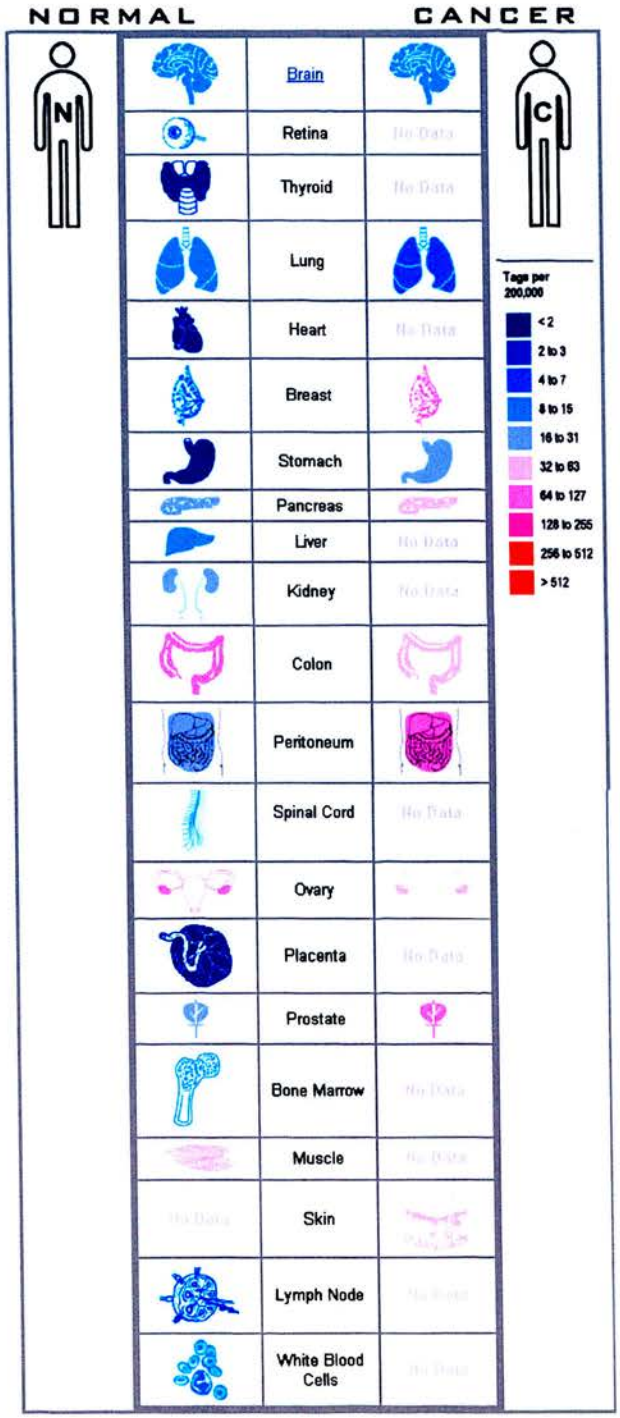


Figure 3.14b Output from SAGE at NCBI for tag representing human Clwd – Normal vs. Cancer expression.

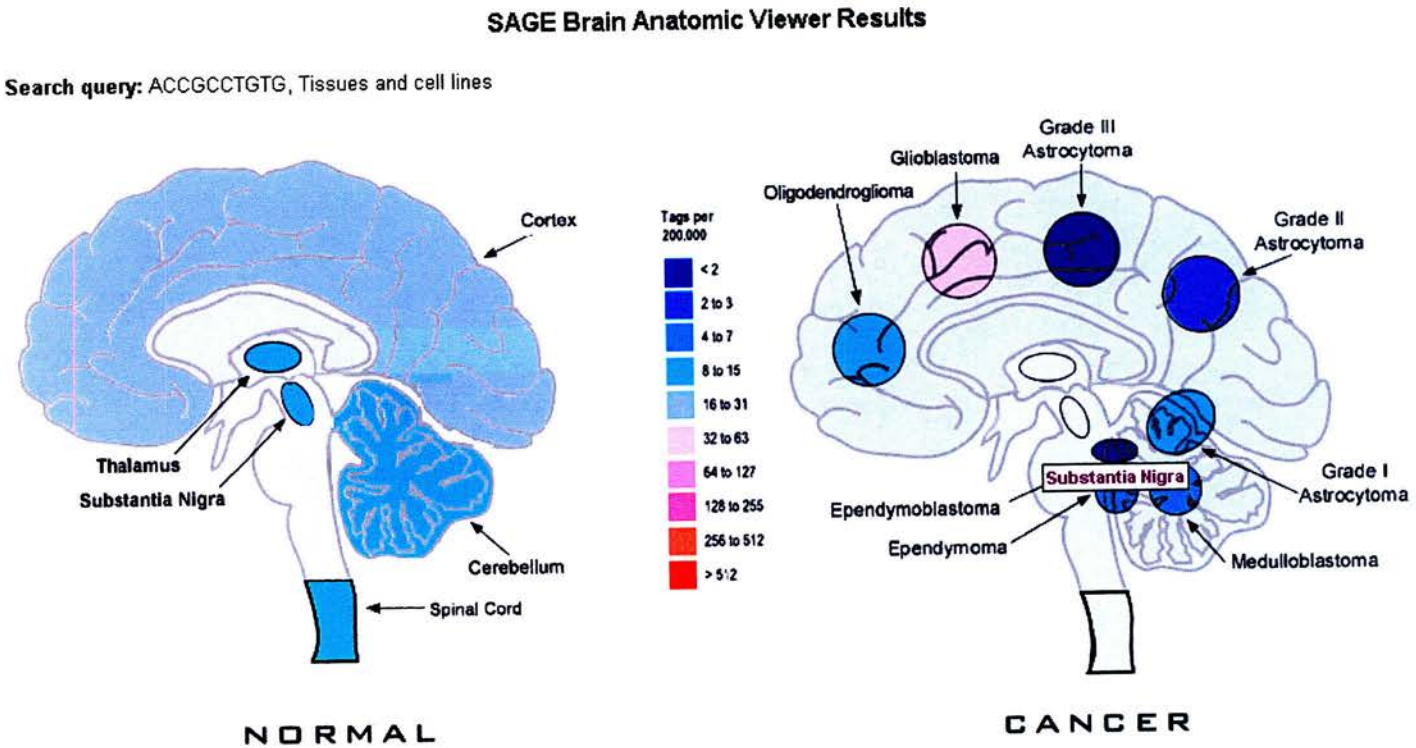


Figure 3.14c Output from SAGE at NCBI for tag representing human Clwd – Normal vs. Cancer expression (expanded view of brain data)

3.4 Discussion

On comparing the sequence of the ESTs that represent Clwd, with the genomic DNA sequence at the telomeric end of mouse chromosome 2, it became clear that Clwd was a slightly unusual gene. It is 1.2kb in length, with 3-5 exons and introns of between 72bp and 209bp in size; given the average mouse intron is 3,888bp in length and the mean number of exons per transcript is 8.4[50], Clwd is unusually small. Does this have any implications for Clwd function? On examining the intron size of a gene with its level of expression it has been found that in mammals at least, introns in highly expressed genes are substantially shorter than those in genes that are expressed at low levels as short introns minimize the slow and expensive process of transcription and splicing[51]. This implies Clwd is a highly expressed gene in addition to a widely expressed gene as suggested by the EST data.

Clwd is alternatively spliced like approximately 41% of all mouse genes[50]. The form of alternative splicing seen in Clwd can be termed length difference alternative splicing, or LDAS as it results in the deletion of a part of the predicted protein sequence. LDAS can create dominant–negative regulators, with the short isoform usually functioning as the regulator[52]. In the case of Clwd this could mean that the short form of the protein, if produced by splice forms II, IV and V, could regulate one produced by I and III and so the relative abundance of the alternatively spliced transcripts in any particular tissue could affect the normal function of Clwd.

There are processed pseudogenes present in the mouse genome that appear highly similar to Clwd. Processed pseudogenes arise through retrotransposition of spliced or partially spliced mRNA into the genome; they are often recognized by the loss of some or all introns relative to other copies of the gene. They sometimes contain all exons, but often have suffered deletions and rearrangements. Over time, pseudogenes tend to accumulate mutations that clearly reveal them to be inactive, such as multiple frameshifts or stop codons.

This has indeed occurred in the case of the Clwd pseudogenes. What is interesting is that the three pseudogenes described seem to represent different stages in the

accumulation of these mutations. The pseudogene on chromosome X is highly similar to splice form I, with only a small deletion and subsequent change in frame late in the sequence distinguishing the ORF from the "real" Clwd. The chromosome 14 pseudogene has an increased number of point mutations while the chromosome 10 pseudogene has a large number of point mutations. It is thus possible to surmise that the retrotransposition event that created the chromosome X pseudogene occurred most recently, with the chromosome 10 pseudogene being the oldest.

Clwd also exhibits an upstream open reading frame. This can dramatically influence the translation of the main ORF. Only 10% of vertebrate mRNAs carry a uORF and this population is largely biased toward proto-oncogenes and genes producing growth factors and growth factor receptors.

An upstream open reading frame can play a major role in the expression of the main protein-producing ORF. One example of this in mammals is the AdoMetDC gene whose 5'UTR contains a six-codon uORF that represses translation of the main ORF in T cells five-fold while in non-lymphoid cells there is little or no effect. Another important aspect to this uORF-mediated translational repression is the gene architecture surrounding the uORF *i.e.* the closer the uORF to the main ORF, it seems, the more efficient the repression[53].

So what does all this mean for Clwd? Although it is by no means decisive, the presence of this uORF in the Clwd mRNA could mean it is a proto-oncogene, or a gene that produces a growth factor or growth factor receptor. The fact that the 5'UTR architecture changes with the different splice forms of Clwd could mean that if this uORF is an inhibitor of Clwd translation, then this inhibition could vary with splice form. In this way the relative abundance of each form in each tissue could affect the abundance and thus function of Clwd gene product.

Using programmes to predict various biological features of Clwd has not been an unqualified success. The subcellular localisation programme results are contradictory and while a motif –based search may offer all the possible ways Clwd could be modified, it is far from certain if these modifications occur *in vivo*. In addition the

lack of well-defined domains and any homology to a known protein makes it all the more difficult to begin to arrive at a possible function for Clwd.

However, given the nature of Clwd as an undescribed and apparent pioneer gene an *in silico* study was the best way to start obtaining information and acquiring ideas regarding the characteristics and function of this new and unusual gene. It revealed where the gene lay, its sequence and protein product. This could be used to design primers, antibodies and other biological tools to begin to characterise this gene. It also identified interesting features that would be worth further investigation such as an upstream open reading frame and the presence of alternative splice forms – are they true biological features of this gene and what is their function, if any? It has also given some clues as to Clwd's function. Does the presence of a uORF suggest a growth factor? Does the apparent up-regulation of expression in some forms of cancer suggest a role as an oncogene?

These bioinformatic tools proved useful in initiating a characterisation of the novel gene Clwd as they provided a starting point for a series of biological experiments to confirm their predictions. These are described in the following chapters.

CHAPTER 4

CHARACTERISATION OF CLWD I:

EXPRESSION OF mRNA AND PROTEIN

4. Characterisation of Clwd I: Expression of mRNA and Protein.

4.1 Introduction

Having used bioinformatics to characterise Clwd as much as possible, I started *in vitro* rather than *in silico* experiments. I decided that my first priority would be to confirm that Clwd is genuinely being transcribed. I would do this by Northern blot. Once this was done, the presence of any alternative splice forms could be ascertained and subsequently the expression of these splice forms in mouse tissue could be analysed by RT-PCR. Given the SAGE results discussed in chapter 3, I also felt it would be important to see if there was a difference in Clwd expression on comparing normal to cancerous tissue.

Having confirmed the presence of Clwd mRNA in tissues, the next goal would be to confirm that any Clwd was translated by designing an antibody against Clwd that could be used in several applications including Western blotting and immunocytochemistry.

4.2 Expression of alternative splice forms of Clwd

4.2.1 Initial confirmation of splice form expression by Northern blotting.

I decided that the best way initially to confirm that Clwd is transcribed was to use a Northern blot. IMAGE clones corresponding to Clwd ESTs were available from HGMP and I used one of these (IMAGE 468225) as the template for making an antisense Clwd probe (and a negative control sense probe). RNA was extracted from various mouse tissues, including heart, spleen, liver and gut and these extracts were run on a formaldehyde gel, before transferring them to a nylon membrane. I labelled the RNA probe with digoxigenin or "DIG" as described in chapter 2 and hybridised with the DIG-labelled Clwd RNA probe. The results are seen in figure 4.1.

Clwd was present in all tissues tested. This confirmed the wide range of expression implied by the EST data. Interestingly, there was also evidence of more than one band in each lane representing different sizes of Clwd mRNA transcript, *i.e.* different splice forms. It appeared that there were three or four splice variants visible in some of the tissues. An approximate idea of the size can be made by using the two rRNA bands (4.712kb and 1.869kb) as a guide but otherwise it is difficult to accurately ascertain the size and thus the identity of these bands.

The Northern blot provided the first concrete evidence that not only did the Clwd gene produce detectable mRNA, but also mRNA that was alternatively spliced. However, I believed it would be difficult to confirm the exact identity of these splice forms by Northern blot alone and so elected to use RT-PCR to identify which splice forms were expressed where.

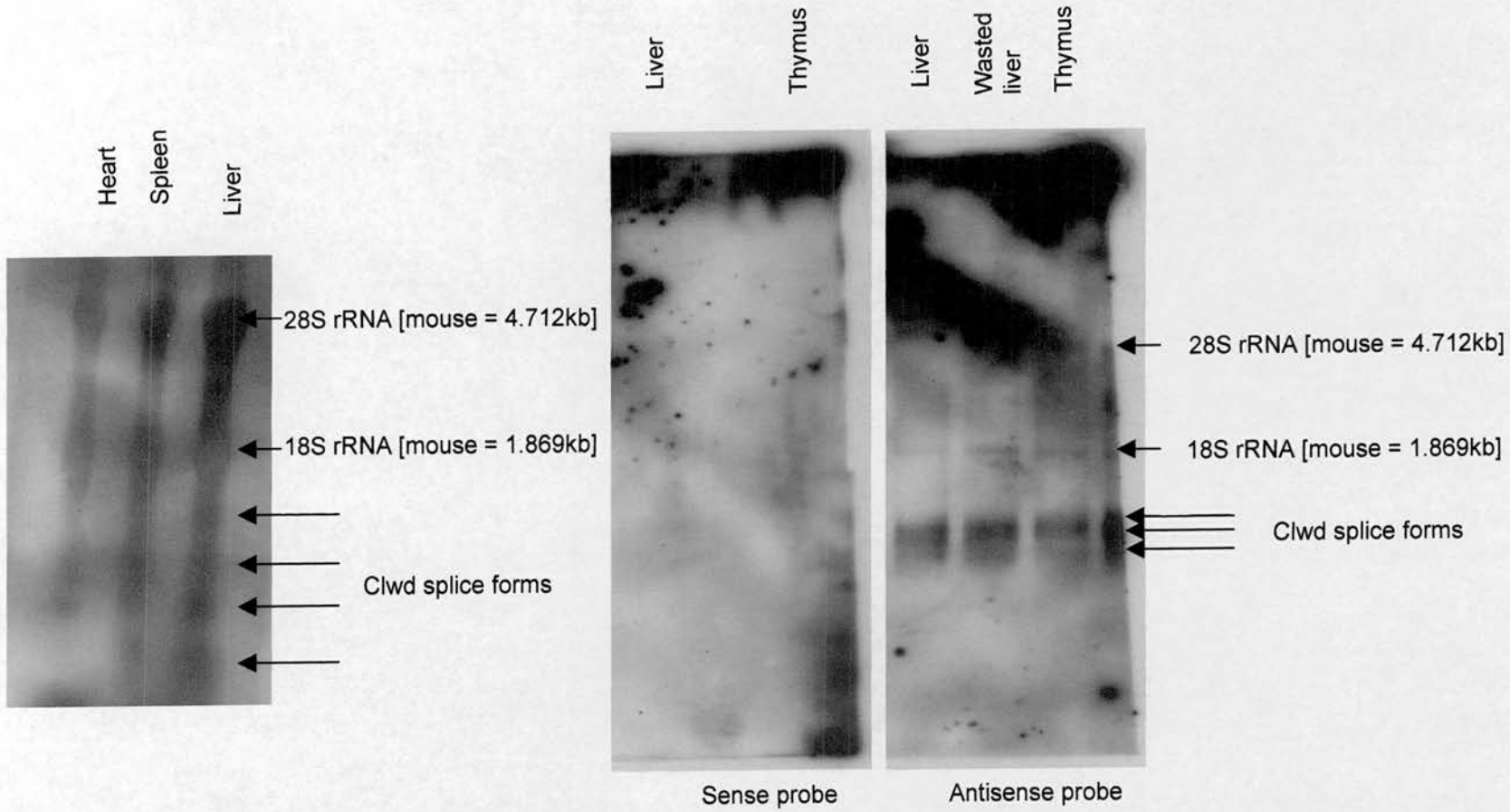


Figure 4.1 Northern blot of various tissues' RNA hybridised with a Clwd-specific RNA probe. These show at least three if not four splice forms hybridising with a Clwd-specific antisense probe in heart, spleen, liver and thymus. A sample of wasted liver RNA is also included here but a comparison of Clwd expression between wild type and wasted mice will be discussed in detail in chapter 6.

4.2.2 Distinguishing expression pattern of Clwd splice forms by RT-PCR

Having confirmed that there was at least some alternative splicing of Clwd in mouse, I wanted to answer a number of questions. Were they as I had predicted from the EST data? Was there any specificity with regard to their expression in various tissues? Could I gain any insight into their relative abundance in tissues? I decided to achieve this by RT-PCR by designing specific primers to distinguish the splice forms.

The primers I used were Clwd A, B, C, D, E, F and "TotalClwd" (forward and reverse) as described in chapter 2. These were designed to maximise the difference in sequence between the splice forms predicted, while the product of the TotalClwd PCR would be predicted to represent total Clwd expression as the primers were in exons 4 and 5 which remain constant through all the splice forms. Where these primers appear in relation to splice forms and genomic DNA are seen in figure 4.2.

The sizes of PCR product produced by each of these pairs of primers were as follows. These primer sequences are described in section 2.2.1.4 in chapter 2.

Position of primers	**Exon 1-2**	**Exon 1-2**	**Intron 1-exon 2**	**Exon 2-3**	**Exon 4-5**
	PCR A-E (bp)	PCR B-E (bp)	PCR C-E (bp)	PCR D-F (bp)	PCR TotalClwd (bp)
Splice form I	-	65	-	120	450
Splice form II	-	65	-	191	450
Splice form III	-	180	-	120	450
Splice form IV	296	180	-	191	450
Splice form V	-	-	156	191	450

While these primers did maximise the difference between the splice forms, I was unable to design primers such that each splice form would give a unique band. I could, however, distinguish splice forms IV and V. Splice form IV was the only form that had a slightly larger first exon with sequence not present in the other splice forms' exon 1 and so a primer designed to this (A) in conjunction with one further

downstream (E) would identify it specifically. Splice form V retains a little more of intron 1 than the others and so primer C was designed specifically to this region, and so should not recognise anything other than splice form V.

It was more difficult to distinguish between splice forms I, II and III as there was no feature unique to these forms that I could use to my advantage. In that case, therefore, the best option was to distinguish between the splice forms that do and do not retain part of intron 1 and likewise for intron 2. In particular I was interested in the retention of intron 2 as this, I had predicted, had a short open reading frame producing a truncated peptide and so it was of interest to see if the splice forms producing this were as abundant as the others.

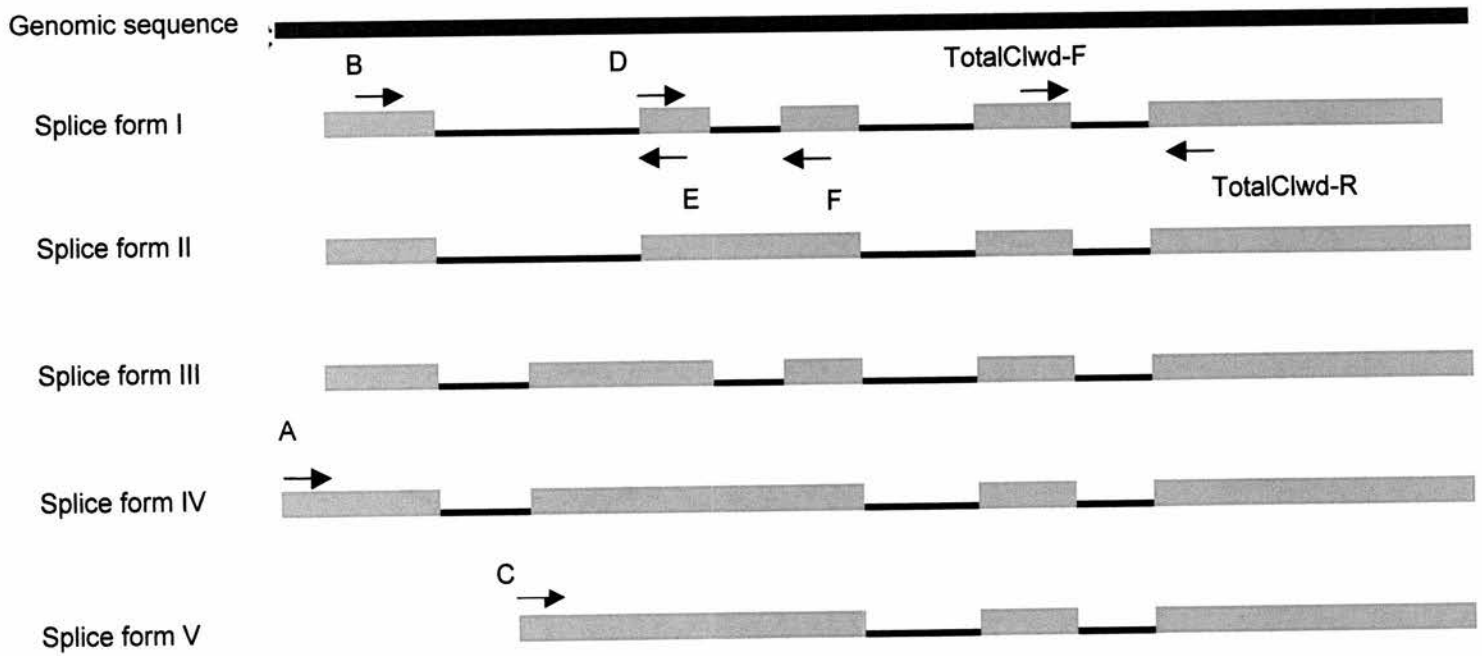


Figure 4.2 Diagram showing alignment of 5 splice forms and situation of PCR primers. Primers A, B, C, D, E, F and TotalClwd were designed in order distinguish the types of isoform present in each tissue examined by RT-PCR. For example the PCR D-F will highlight the retention/splicing of intron 2 in each tissue. The PCR bands predicted to result from each PCR can be seen in the table on page 98.

The tissues used in the RT-PCR were from a 17 day old wild type mouse from the wasted stock (see chapter 2 for strain details). The RNA was extracted and cDNA prepared as described in chapter 2. The primers described were then used to amplify regions of the cDNA to ascertain the splice forms present. Two types of control were included. During the cDNA synthesis, for each RNA sample that was reverse transcribed, another RNA sample from the same tissue was processed, except without the reverse transcriptase that is vital to produce cDNA. This was necessary to confirm that there was no genomic DNA from the tissues carried over during the RNA preparation. If this control was negative for each sample, I would know that all PCR products seen in the reverse transcribed samples were derived from mRNA rather than contaminating DNA. Bearing in mind that there were three pseudogenes identified for Clwd (chapter 3) this control was particularly important.

The second was a PCR control, where for each set of primers (and thus master mix) one sample to be amplified by PCR contained no DNA template. This would ensure, given a negative result for this sample, that there were no contaminating PCR products or other sources of DNA template for the PCR present in the components of the reaction.

The results of the RT-PCR are shown in figure 4.3

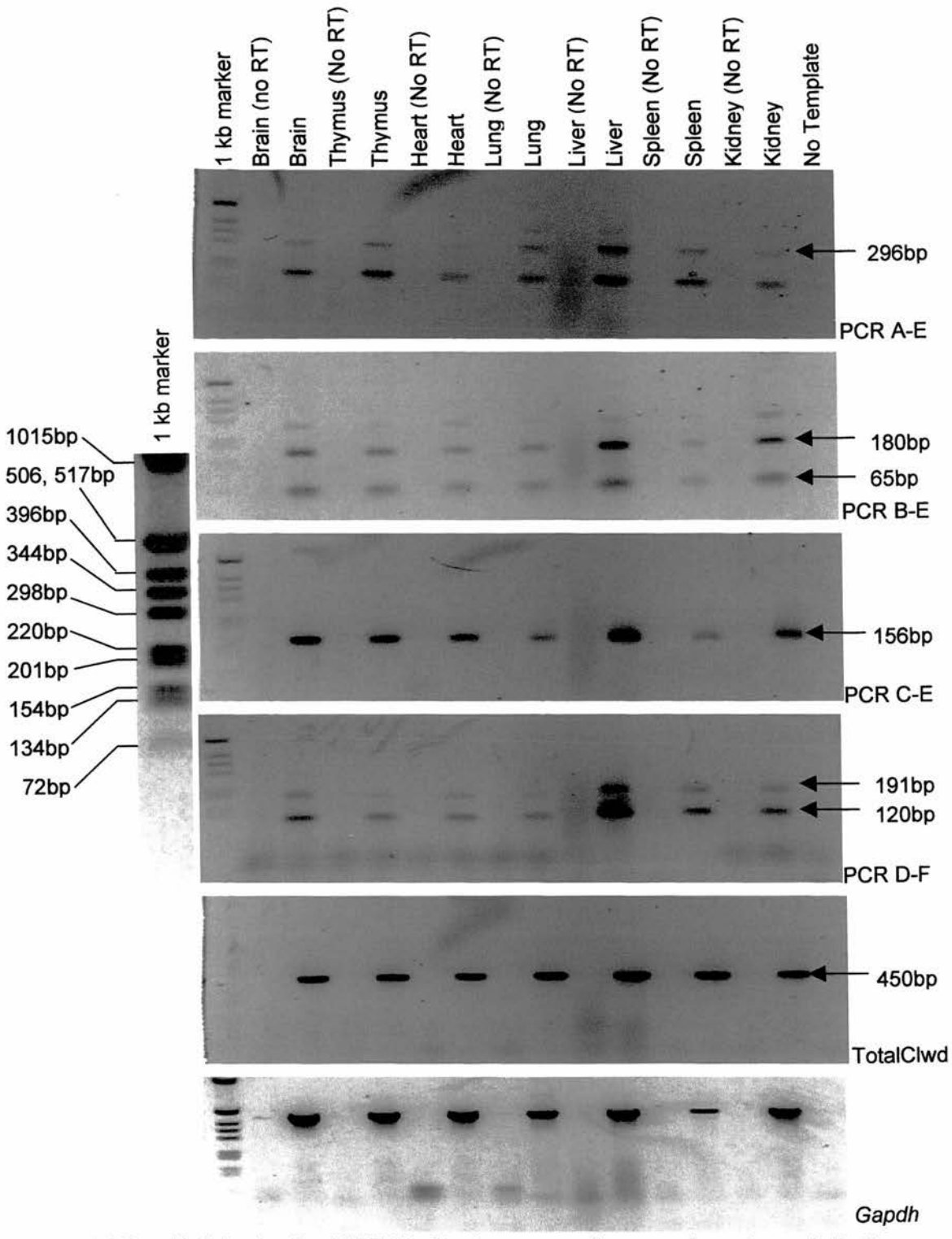


Figure 4.3 Results following the RT-PCR of various mouse tissues using primers A, B, C, D, E, F and TotalClwd. The sequences of these primers can be found in section 2.2.1.4 of chapter 2. The sizes of PCR product denote the types of splice form present in each tissue (see table on page 98 for information). Here *Gapdh* primers were used to amplify the housekeeping gene *Gapdh* as a control for the RNA and thus cDNA levels in the samples used

There were a number of points to be noted about this experiment.
Firstly, the lack of bands in the negative controls, both the RT negative and PCR negative control, suggested that the bands seen in the lanes representing the tissue samples did truly represent the amplified portions of mRNA, and not any form of contamination.

Secondly, it can also be seen that there were extra bands in both PCR A-E and PCR B-E that were not predicted to be produced by the 5 known splice forms. When these were examined more closely it could be seen that the extra band in PCR B-E is larger than the predicted bands of 180bp and 65bp in length. Given that the band is quite faint and that there are two products produced by this PCR it is most likely that this band is due to the formation of a heteroduplex. This is where one strand of one product anneals with one strand of another at their compatible 5' and 3' ends with a loop of their non-compatible sequence in the middle. Bands such as these are retarded in their progress through the gel and so they appear larger than the two products of which they are comprised.
There were also extra bands in PCR A-E which should distinguish splice form IV from the others by using a primer in exon 1 which is specific for a sequence at the beginning of this exon not seen in the other splice forms.
Here again, the larger unpredicted band appears to be a heteroduplex of the two other products. The smaller unexplained product could represent a splice form that does not retain intron 1 at all - giving a band of 178bp. There could be two possible causes of this; either there was another undescribed splice form with the larger exon 1 similar splice form IV and a non-retained first intron, or splice form IV did not actually have an extended exon 1 but rather all the splice forms had this, and so PCR A-E produces the same result as PCR B-E in fact.
As mentioned in chapter 3 (see figure 3.2a) the sequences of splice forms types I-V used to design the primers in this experiment are taken from individual clones. Therefore the 5' end of these could be different from that described. This seems to be the case when considering the results of PCR A-E *i.e.* exon 1 is longer than that seen in splice forms I-III. Given this possibility, I have not considered the results of PCR A-E to be indicative of the presence/absence of splice form IV.

These results therefore, rather than distinguishing one splice form from another, were really useful in distinguishing the different types of isoform from one another *i.e.* there are Clwd splice forms that partially retain and fully retain intron 1 in their final processed mRNA, and that there are splice forms that do and do not retain intron 2 in the final processed mRNA. Splice form V could be distinguished from the others.

In addition to the primers described above I also subjected these samples to PCR using a set of primers specific to *Gapdh* (Glyceraldehyde-3-phosphate dehydrogenase). This is a gene that is strongly expressed, at all times and in all cells as it is required for their basic metabolic function. As such it is known as a "housekeeping gene" and so could be used as a constant against which the intensity of the product produced by the various PCRs of the Clwd splice forms could be measured.

As can be seen from the results of this experiment above, the intensity of the *Gapdh* band is not constant when compared from tissue to tissue. This was expected as it is extremely difficult to measure accurately the exact amount of RNA that is subjected to cDNA synthesis and there can be significant variation in the efficiency of reverse transcription. As a result, I decided to make an approximate measure of the amount of product produced by expressing it as a fraction of the *Gapdh* product produced. I did this by using Scion Image software (Scion Corp) to measure the intensity of the bands produced by PCR and expressing the products of the Clwd PCR as a fraction of the *Gapdh* product intensity. As can be imagined this is a crude measurement of the level of expression of the Clwd splice forms, but would allow me at least to get a rough idea of relative abundance of the splice forms. Given the results of the RT-PCR above, I believed the most informative way of looking at these results would be to compare the abundance of the two products of the PCR B-E which would represent the partially retained intron 1 (180bp) and the non-retained intron 1 (65bp), and the two products of PCR D-F which would represent the retained intron 2 (191bp) and the non-retained intron 2 (120bp).
A graph of the results is shown in figure 4.4

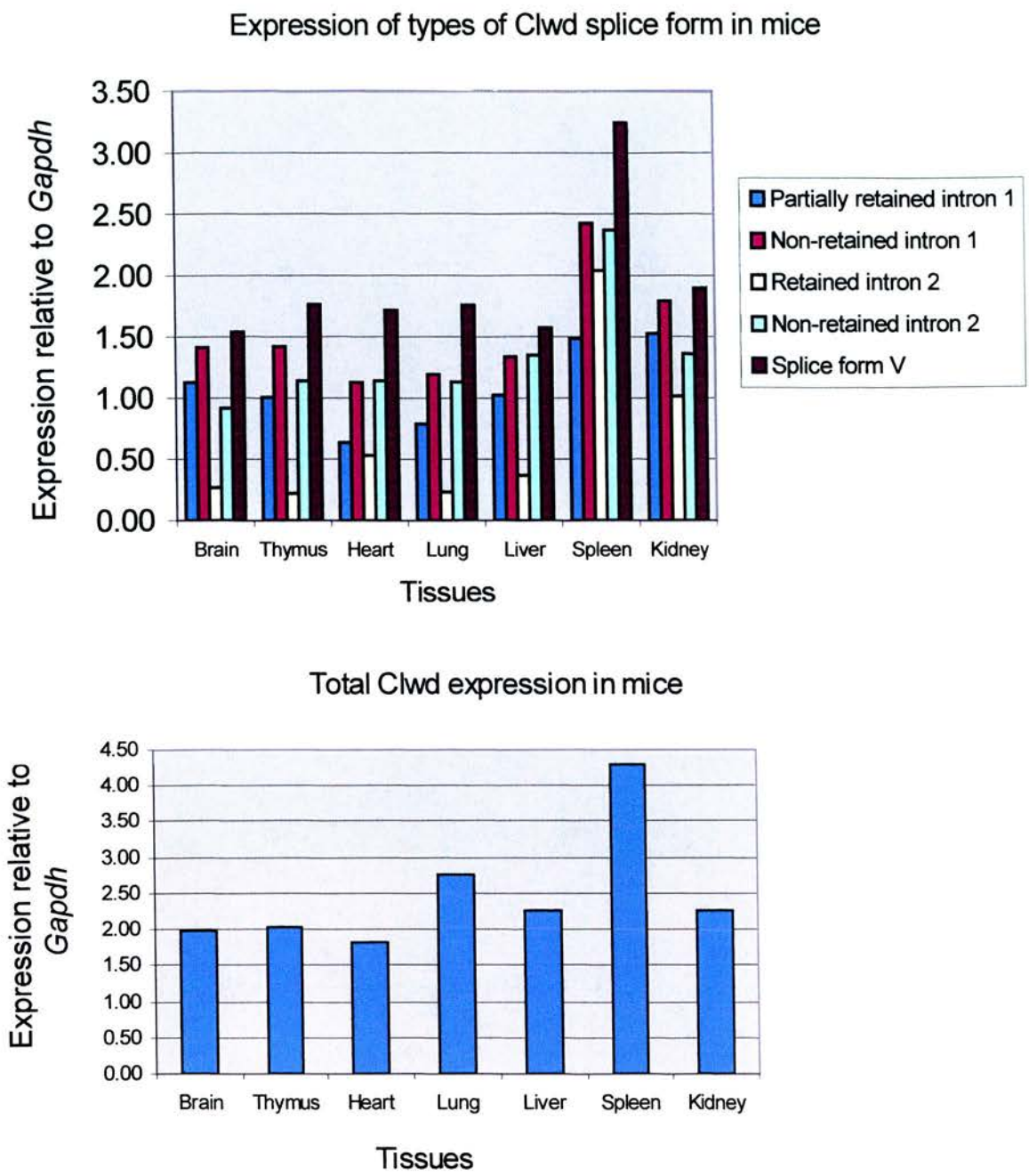


Figure 4.4 Graph showing relative expression of each splice form of Clwd to that of *Gapdh*

These results showed that there seems to be the following pattern of expression in each of the tissues –partial retention of intron 1 < non-retention of intron 1, while for intron 2 *i.e.* retention of intron 2 << non-retention of intron 2, while there appears to be strong expression of splice form V.

What must be borne in mind however, are the limitations of this particular method of judging expression levels, in particular where two or more products are produced by the same primer set. It can often be seen that when this is the case that the smaller product is always more abundant than the larger as the reaction will preferentially amplify the shorter product. It is worth noting however that the levels of splice forms retaining intron 2 are greatly increased in the spleen. There does appear to be a general increase in Clwd expression in the spleen, but specifically on comparing the levels of splice forms retaining the second intron and not retaining it there seems to be a far greater proportion of retained 2^{nd} intron. This is especially noticeable when looking at the retention of intron 1 in spleen, which seems hardly changed at all compared to the other tissues.

4.2.3 Matched tumour/normal expression array

In chapter 3 I described a database containing SAGE data from normal tissues and tumours which, when queried with the Clwd tag sequence, suggested that Clwd might exhibit increased expression in various cancers. In an attempt to corroborate this with some *in vitro* evidence, I used a matched tumour/normal expression array available from Clontech. This array consists of SMART™ -amplified cDNA from 68 tumour and corresponding normal tissues from individual patients spotted onto a nylon membrane Clinical information for each individual sample can be seen in section 2.A.3.

An IMAGE clone known to represent the human form of Clwd was used as a template to produce a DIG-labelled Clwd cDNA probe by random primed labelling. The array was then be incubated in a hybridisation solution containing the DIG-labelled cDNA probe as described for a Northern blot in section 2.2.2.7 and 2.2.2.8. The result can be seen in figure 4.5a with analysis in figure 4.5b.

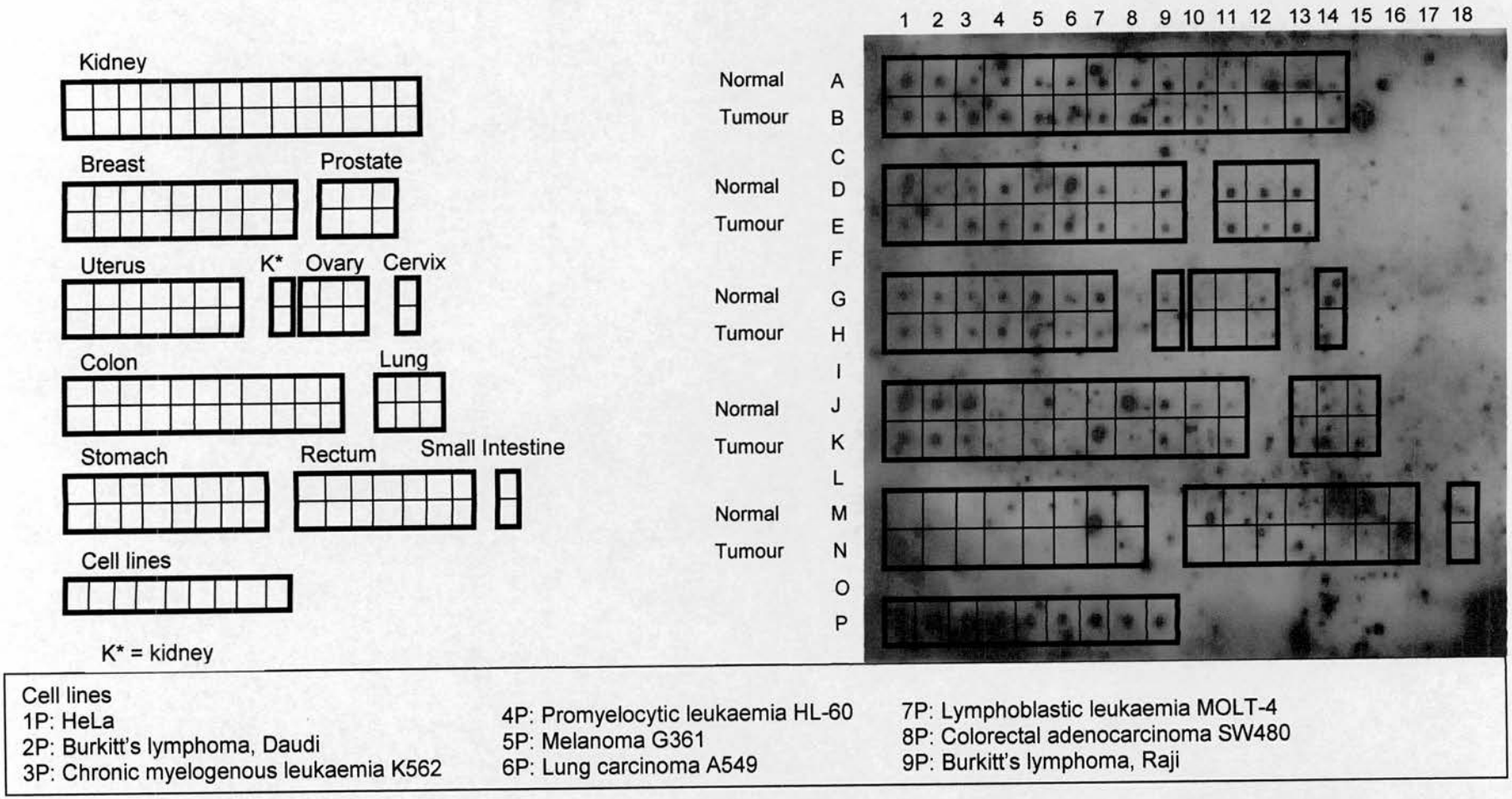


Figure 4.5a Matched tumour/normal expression array hybridised with Clwd-specific cDNA probe. The array consists of cDNA samples from normal and tumour tissue from individual patients. The patients are grouped by the tissue in which the tumour arose. For example samples 1A and 1B are cDNA from one patient with a kidney tumour from normal and cancerous tissue respectively.

A

Tissue	Clwd expression	Number of patients showing change in expression	Number of patients showing no change in expression
Kidney	Yes	6	8
Breast	Yes	1	8
Prostate	Yes	1	2
Uterus	Yes	2	5
Ovary	No	-	-
Cervix	Yes	0	1
Colon	Yes	5	6
Lung	Yes	0	3
Stomach	No	-	-
Rectum	No	-	-
Small intestine	No	-	-

B

	Cancer	Metastasis?	Patient age and sex	Change
3 A/B	Kidney renal cell carcinoma	None seen	44yrs female	Increased Clwd in tumour
4 A/B	Kidney renal cell carcinoma	Metastasis to bone	57yrs male	Increased Clwd in tumour
5 A/B	Kidney renal oncocytoma	Not available	73yrs male	Increased Clwd in tumour
9 A/B	Kidney renal cell carcinoma	None seen	50yrs female	Decreased Clwd in tumour
12 A/B	Kidney renal cell carcinoma	None seen	47yrs male	Decreased Clwd in tumour
13 A/B	Kidney renal cell carcinoma	None seen	56yrs male	Decreased Clwd in tumour
3 D/E	Breast infiltrating duct carcinoma	Metastasis to lymph nodes	45yrs female	Increased Clwd in tumour
1 J/K	Colon benign tumour	Not available	81yrs male	Decreased Clwd in tumour
2 J/K	Colon adenocarcinoma	Metastasis to liver	68yrs male	Decreased Clwd in tumour
3 J/K	Colon adenocarcinoma NOS*	None seen	64yrs male	Decreased Clwd in tumour
7 J/K	Colon adenocarcinoma NOS*	None seen	69yrs female	Increased Clwd in tumour
8 J/K	Colon adenocarcinoma NOS*	Metastasis to lymph nodes	35yrs female	Decreased Clwd in tumour

*NOS – Not Otherwise Specified

Figure 4.5b Tables showing analysis of matched normal/tumour expression array data. Table A summarises Clwd expression on the array *i.e.* the presence or absence of Clwd and/or any change in Clwd expression on comparison of the normal and tumour sample from each patient. Table B shows the in depth detail of the patients for whom a change in Clwd expression was recorded. These changes were estimated by eye from the appearance of the array as seen in figure 4.5a.

Although it is difficult to say with certainty given the background, it is worth noting that there appears to be little if any Clwd expression in the ovary, stomach and rectum. At the very least there appears to be far higher expression elsewhere, particularly the kidney, breast, prostate, uterus and colon. As for changes in Clwd expression with the onset of a tumour, this seen in the kidney and colon tumours primarily, with both increases and decreases noted.

4.3 Expression of CLWD protein

4.3.1 Introduction

On looking at the splice forms produced by Clwd it is clear to see that there are several categories of splice form, detectable by RT-PCR. To begin to explore the function of Clwd, the next priority was to ascertain if there was a protein product produced. I decided that the first step required to look at the Clwd protein would be to make an anti-Clwd antibody.

4.3.2 Design and production of an anti-peptide antibody against Clwd.

Anti-peptide antibodies had been prepared and used successfully against various proteins by our group prior to my beginning this study. This, in addition to the fact that their production is far less labour-intensive than the production of bacterially-expressed proteins, was the main reason for choosing this method. The first step to making this antibody was the design of a suitable peptide. To do this I had to take several things into account. Firstly I wanted an antibody that would recognise both human and mouse Clwd and so the peptide would have to come from a region where there is a high degree of conservation between the two proteins. Secondly in order for the peptide to be sufficiently antigenic *i.e.* for it to raise a good antibody response in the sheep it would need to be in a region of the protein that would be exposed to the immune system *i.e.* on the surface of the protein. Given these requirements I used the programme PIX at HGMP to assess the antigenicity along the length of the putative Clwd protein in both mouse and human to select the best candidate sequence.

```
Human: MAAIPSSGSLVATHDYYRRRLGSTSSNSSCSSTECPGEAIPHPPGLPKAD
Mouse: MAAIPSSGSLVATHDYYRRRLGSSSSSSSGGSAEYPGDAVLQSPGLPKAD

Human: PGHWWASFFFGKSTLPFMATVLESAEHSEPPQASSSMTACGLARDAPRKQ
Mouse: PGHWWASFFFGKSTLPFMTTVLESPERSAESPQVSRSPMTCGLTPETMKQ

Human: PGGQSSTASAGPPS
Mouse: QPVIHSGQTNPRDLS
```

The amino acids underlined in the human and mouse Clwd sequences above are the most antigenic portions of the protein *i.e.* they represent the regions that are most hydrophilic and therefore most likely to lie on the surface of a correctly folded protein. As can be seen here the first antigenic site in both human and mouse Clwd are identical and so this was the peptide that I chose in order to raise an antibody. This peptide was chemically synthesised with a cysteine added to the N-terminus to

allow for the addition of keyhole limpet haemocyanin (KLH) which would increase the immunogenicity of the peptide. The peptide was then used to immunise a sheep as described in chapter 2 and I received a sample of pre-immune serum (taken from the animal before immunisation) and three samples of post-immune serum taken at one month intervals.

To begin to assess the efficiency with which this serum could act as an anti-Clwd antibody, I first needed to prove that there were antibodies in the serum that would recognise the peptide that was used to immunise the animal. I did this by carrying out an ELISA (Enzyme-Linked ImmunoSorbent Assay). This ELISA involved linking the peptide to wells of an immunosorbent plate, adding the serum and then an anti-sheep secondary antibody that was conjugated to horseradish peroxidase. This enzyme was then used to effect a colour change on a substrate (OPD – O-PhenyleneDiamine) which would allow me to assess the function of any anti-Clwd antibodies present in the serum.
The results of this ELISA are shown in the form of a graph of absorbance at 450nm versus the $\log_{10}$ of the concentration of the peptide (1-5000ng/ml) for the pre-immune serum and the first two samples of post-immune serum. What can be seen is that while the absorbance at various concentrations of pre-immune serum remained at a low-level, with increasing peptide and serum concentration the absorbance of the post-immune sera increased. Thus the post-immune serum was reacting to the peptide and so did contain anti-Clwd antibodies.

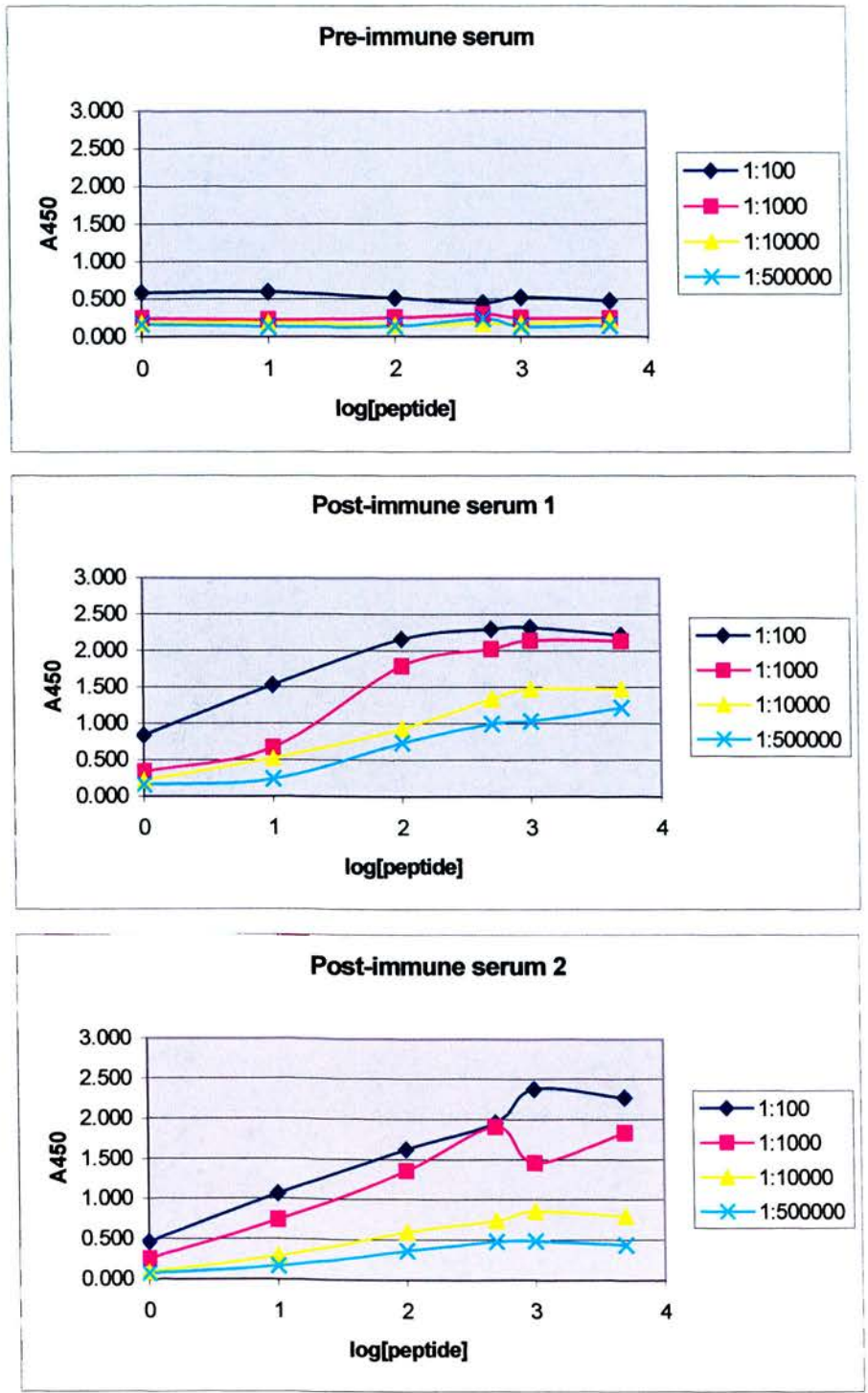
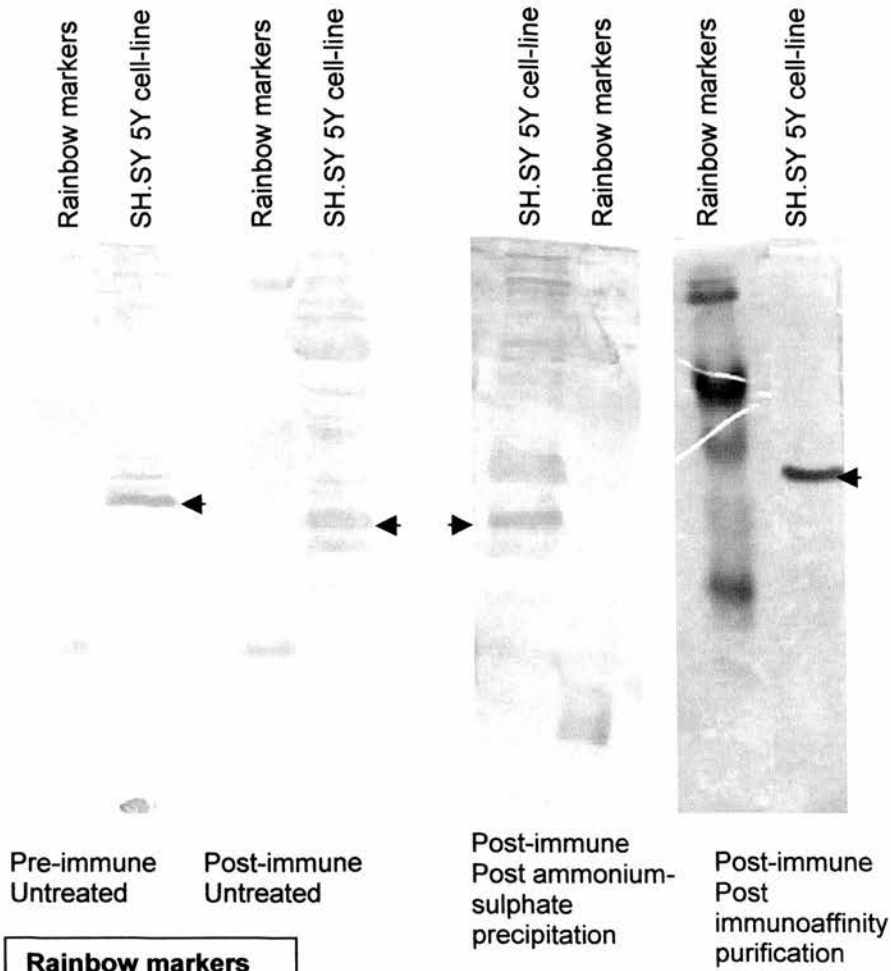


Figure 4.6 ELISA showing the positive reaction of the anti-peptide antibody to the peptide representing Clwd

4.3.3 Purification of antibody

The next step was to purify the post-immune serum in order to maximise the concentration of these antibodies and minimise the presence of the various other components of the sheep serum including other antibodies. As can be seen in figure 4.7, using the non-purified serum in a western blot of a human cell line resulted in a large number of non-specific bands. The first step in purification was an ammonium sulphate precipitation step to yield an immunoglobulin-enriched serum. This reduced the number of bands seen on a western but further purification was required. This was achieved by using an immunoaffinity purification step. I covalently attached the peptide to a SulfoLink column and used this to extract and then elute that component of the serum that would bind the peptide. This resulted in a single band when used on a Western blot of a human neuroblastoma cell line (SH.SY 5Y). I had now purified the serum to a form that could be used successfully on a Western blot.

It is important to note that this band albeit single and apparently specific was approximately 28kD in size while, from the sequence of Clwd alone, the size of the protein would be calculated at approximately 12kD. This would suggest that Clwd was post-translationally modified which increased its size.



Rainbow markers	
Blue	202.0kD
Magenta	121.0kD
Green	79.0kD
Violet	41.0kD
Orange	31.6kD
Red	17.8kD
Blue	8.2kD

Figure 4.7 Western blots detailing the increase in specificity of the anti-Clwd antibody following rounds of purification. Arrowhead on each Western blot indicates the position of the final strongly immunoreactive band as the serum is purified. Each step in purification decreases the background level until only one band remains in the final blot. This band is seen at 28kD rather than at 12kD as would be expected for the Clwd protein. This may indicate post-translational modification of Clwd takes place *in vivo*

4.3.4 Expression in tissues

Having purified the serum so that it could be used as an anti-Clwd antibody, I then tested a range of mouse tissues to see where the Clwd protein was produced. The results are shown in figure 4.8 below. What can be seen here is that Clwd does appear to have an increased size compared to that which would be predicted from its amino acid sequence alone. In addition some tissues, most notably spleen, but also liver, exhibited a doublet. This could be due to a form of post-translational modification that is dynamic in nature and so in a tissue one would find both the modified and non-modified forms – resulting in a slightly different size of protein on a Western blot. An example of this kind of modification would be phosphorylation or O-GlcNAcylation.

From this and other Western blots it became apparent that Clwd is as widely expressed as protein as it is mRNA, with every tissue examined showing Clwd expression.

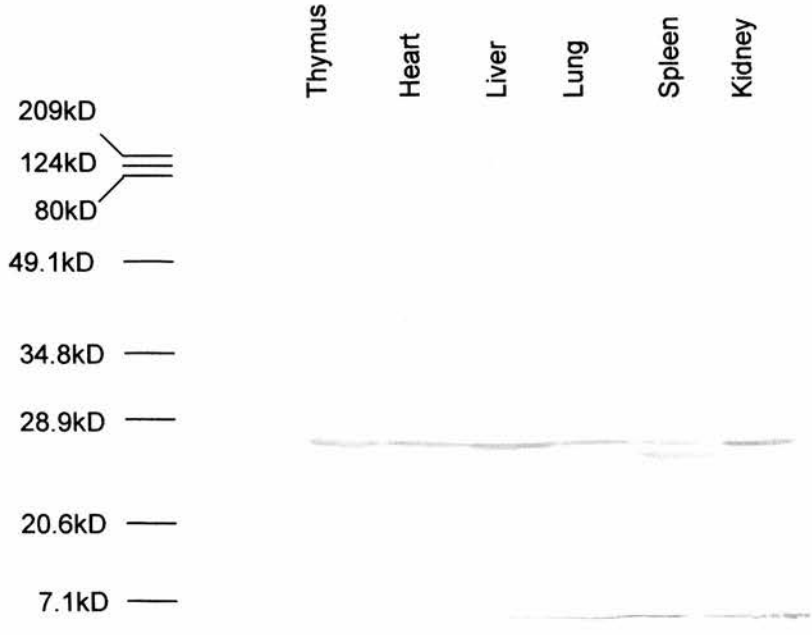


Figure 4.8 Western blot of mouse tissues using anti-Clwd antibody. Clwd expression could be detected in all tissues examined. A doublet was observed in the liver and spleen sample. This could be due to different phosphorylation states of the Clwd protein *in vivo*

4.4 Discussion

In this chapter I have described some of the first experiments I conducted on the novel gene Clwd. Taking the information I had gleaned from the sequence of the gene and open reading frame into consideration allowed me to design probes/primers and a peptide/antibody to confirm the presence of Clwd as an expressed mRNA and protein respectively in a wide range of mouse tissues.

The RT-PCR results were very interesting. They confirmed some of the splice forms that I had seen represented as ESTs and seemed to suggest that overall the balance of splice forms is maintained from tissue to tissue. I also described some semi-quantitative analysis where I compared the intensity of the products produced by the various PCR primers with that produced by *Gapdh*-specific PCR primers to get an idea of the levels of Clwd seen in various tissues. Firstly, Clwd appeared to be expressed to a similar degree in all the tissues examined, with particularly high expression in the spleen. Secondly, of the types of splice form expressed it appeared that those that do not retain the first and second intron are more highly expressed than those that do. In the spleen however it also seemed that there was a larger proportion of Clwd transcripts present with the second intron retained. This could suggest a reduction in the level of Clwd's full ORF if this splice form acts as a repressor as discussed in chapter 3.

It must be pointed out however that this analysis is a crude method of obtaining relative expression, as PCR is by no means a good way of quantitating expression. Given these interesting suggestions though, and had I more time, the use of real time PCR would be an excellent way of confirming these results.

Given the results obtained by SAGE, I also chose to look at total Clwd expression in cancer as a preliminary examination of its role in tumourigenesis. By using the matched normal/tumour expression array and a Clwd probe to hybridise it, it was possible to determine the expression of Clwd cDNA in the tissues included on the panel. In some of the tumours there were changes in Clwd expression. Both

increased and decreased Clwd expression was noted with changes predominantly seen in kidney and colon cancers. These included cancers that subsequently metastasised and those where no metastasis was seen. It was impossible, therefore, to draw any general conclusions from the results (*e.g.* in all colon cancers, Clwd is increased). However, the changes in Clwd expression whether peculiar to the tissue studied or to some tumourigenic state would be worth pursuing, particularly by immunohistochemistry, even if that change represents a secondary effect of the cancer on Clwd expression.

The antibody raised against Clwd appears to work well in that it identifies a single specific band on a Western blot. This band is not of the size predicted by the amino acid sequence, but could be accounted for by post-translational modification.

The experiments described here, allowed me to confidently say that Clwd was a gene that produced mRNA and protein that appeared to be widely expressed as I had predicted using various bioinformatic programmes. They had also provided me with one of the most valuable tools with which I could further elucidate the function of Clwd, an antibody.

CHAPTER 5

CHARACTERISATION OF CLWD II:

LOCALISATION AND FUNCTION

5. Characterisation of Clwd II – Localisation and Function.

5.1 Introduction

Having confirmed that the novel gene Clwd is both transcribed and translated, the next question to address was the role of Clwd in cells. To accomplish this it was important to examine the subcellular localisation of the Clwd protein. If it was possible to detect Clwd within the cell, ideally localised to a discrete organelle, this would help in determining a potential function. The subcellular localisation experiments were done by immunocytochemistry in two cell lines, the human cervical carcinoma line, HeLa and the mouse embryonic fibroblast cell line, NIH/3T3.

The establishment of the correct conditions for successful immunocytochemistry would also be valuable for the next experiment to directly address Clwd's function *i.e.*, the use of RNAi in cell culture in order to silence Clwd and ascertain the effect this has on the cell.

RNAi or RNA interference is the term used to refer to the method of silencing genes post-transcriptionally using double stranded or dsRNA. This form of gene silencing had been used successfully as a powerful tool in the study of *C. elegans* genes for some time before the method became tractable in mammalian cell culture. Studies have shown that this method of gene silencing is mediated by the generation of small interfering RNAs produced from the dsRNA molecules by the enzyme Dicer. These siRNAs subsequently target the specific degradation of the mRNA to which they are specific by an enzyme complex[54].

The use of RNAi in mammalian cell culture was first described by Elbashir *et al* in Nature[55]. It was discovered that while long stretches of dsRNA can function to silence genes in *C. elegans*, these trigger an antiviral defence pathway in mammalian cells involving the activation of PKR, a kinase that phosphorylates EIF2a, which responds by shutting down all translational activity[54]. siRNAs used to silence genes in mammalian cell culture take the form of base-paired 21 nucleotide double stranded RNAs with 3' overhanging ends. These siRNAs avoid triggering the

activation of PKR due to their small size and are presumed to be incorporated into the RNAi pathway by mimicking the products of the Dicer enzyme, which catalyses the initiation step of RNAi.

There has been some controversy regarding the specificity of these siRNA molecules for their target.

One study in particular cast doubt on this specificity when they showed transcript profiles which had siRNA-specific rather than target-specific signatures and the direct silencing of non-targeted genes containing as few as eleven contiguous nucleotides of identity to the siRNA[56].

However there have been two further reports on the specificity of siRNAs. One study used DNA microarrays to profile global gene expression upon the application of siRNA to silence an exogenous gene. It was found there was no alteration in expression of any gene other than the one for which the siRNA was specific[57]. The second report was more qualified in its conclusion. It stated that when strict rules were applied to the design of the siRNAs, and the conditions for the transfection were optimised, the technique can be highly specific, *i.e.* the global expression signatures of different siRNAs against the same target correlate very closely, while those for different genes do not. However if a high concentration of siRNA was used (100nM) or there was some homology in the sequence designed and unwanted targets (>15nt in 21nt) then there can be non-specific silencing[58]. Therefore the recommendations of this and other papers were taken into consideration when designing and carrying out the RNAi experiments. These will be described in section 5.3

5.2 Subcellular localisation of Clwd

5.2.1 Fluorescent labelling of Clwd in cells

In order to assess where Clwd was expressed within the cell, several immunocytochemistry experiments were conducted. The first experiment involved testing if the anti-Clwd antibody could be used successfully in this type of protocol and assessing the best method of fixing and permeabilising the cells to the antibody. HeLa cells were grown on coverslips in 24-well plates for one to two days, until they had adhered to the coverslip and were growing and dividing as normal. The cells were then washed and fixed using a variety of methods. These included using ice-cold methanol, an ice-cold 1:1 mix of methanol and acetone or 4% paraformaldehyde. While methanol-based fixatives work by precipitating the proteins in a cell such that diffuse protein may be lost, paraformaldehyde preserves the native structure. With methanol-based fixatives, the cells are also permeabilised by the treatment, but using paraformaldehyde means an extra permeabilisation step is required. I used either triton X-100 or digitonin. Digitonin differs from triton in that while it disrupts the plasmalemma, the intracellular membranes including the nuclear envelope are left intact. Once fixed and permeabilised the cells were then incubated with a blocking buffer containing Tween®-20 (a non-ionic detergent used for disrupting unwanted lipid-protein interactions), BSA and serum from the animal in which the secondary antibody was raised. In this case the secondary antibody for the anti-Clwd primary antibody was a donkey anti-sheep antibody and so the serum used was donkey serum. This blocking buffer prevented the non-specific adsorption of the immunological reagents. The cells were then incubated in primary antibody and secondary antibody, were counterstained with DAPI, a nuclear stain, and mounted on a slide for viewing as described in chapter 2.

In this experiment the following negative controls were used. Firstly the cells were treated with secondary antibody but not primary antibody to ensure that the secondary antibody gave a signal only when it was bound to the primary antibody. Secondly I preincubated the primary antibody with 1ng of the peptide against which

it was raised. If the antibody was truly specific to Clwd and indicated Clwd's presence in a cell, then the addition of the peptide should reduce this signal by blocking those sites on the anti-Clwd antibody that would recognise Clwd.

As can be seen in figure 5.1 it appears that the methanol/acetone method of fixation minimised the signal in the negative controls while maximising that of the positive control and so this was the method used for future experiments. On examining the staining of the cells it appears that the antibody stains the cytoplasm with the strongest signal around the nucleus. The lower panels show the MeOH-acetone fixed HeLa cells with the DAPI counterstain highlighting the staining around the nucleus. Given that there is also staining in the cytoplasm, it is less likely that Clwd was an integral part of the nuclear membrane itself, but rather that its expression while generally cytoplasmic, is concentrated at the interface between the nucleus and the cytoplasm perhaps to perform a particular function.

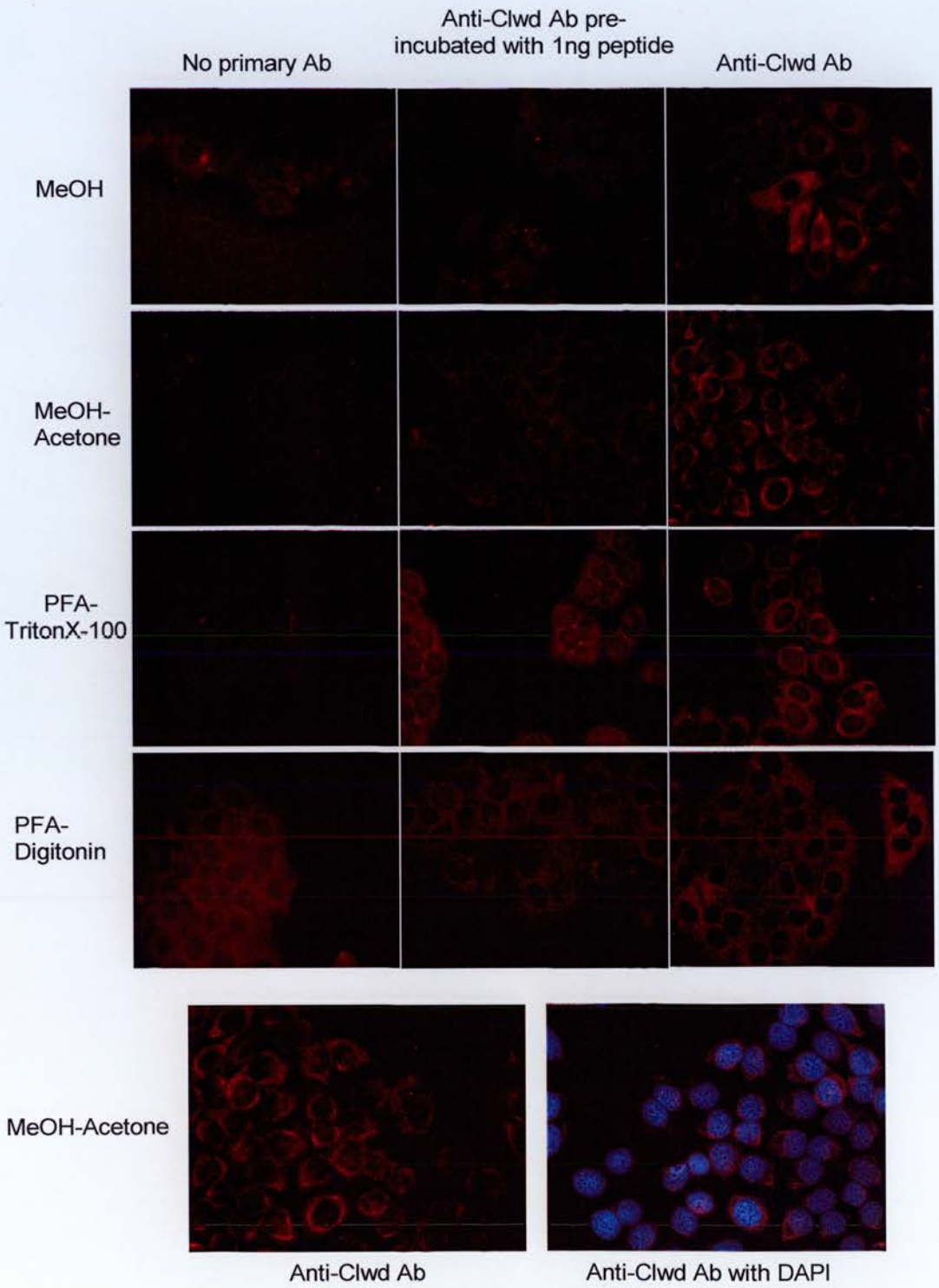


Figure 5.1 Anti-Clwd antibody staining of HeLa cells fixed by a variety of methods

5.2.2 Co-localisation with organelle markers

Following the initial immunofluorescence experiment results, I conducted several double-labelling experiments to determine if Clwd co-localised with a known protein, of specific expression pattern. These experiments were conducted in NIH/3T3 cells, a mouse fibroblast cell line, rather than the human cell line HeLa used in the previous experiments. I first used an organelle sampler kit (BD Transduction Laboratories) that contained antibodies against a variety of proteins that were markers of particular organelles. These antibodies and their associated organelle are listed below:

Protein	**Organelle**
Annexin II	Plasmalemma
β–Catenin	Zonula Adherens
Caveolin 1	Caveolae
GM130	Golgi
Paxillin	Focal Contacts

While these antibodies were raised in mice, the anti Clwd antibody was raised in sheep. This meant that if the anti-mouse antibody and anti-sheep antibody were labelled with different fluorescent probes, both Clwd and the marker of interest could be detected in the same set of cells and the two images merged to determine if they were co-localised. The organelle markers were stained green, Clwd stained red and where these two colours are co-localised in the merged picture the staining appears yellow. The most promising marker, in terms of co-localisation with Clwd is Paxillin, as seen in figure 5.2.

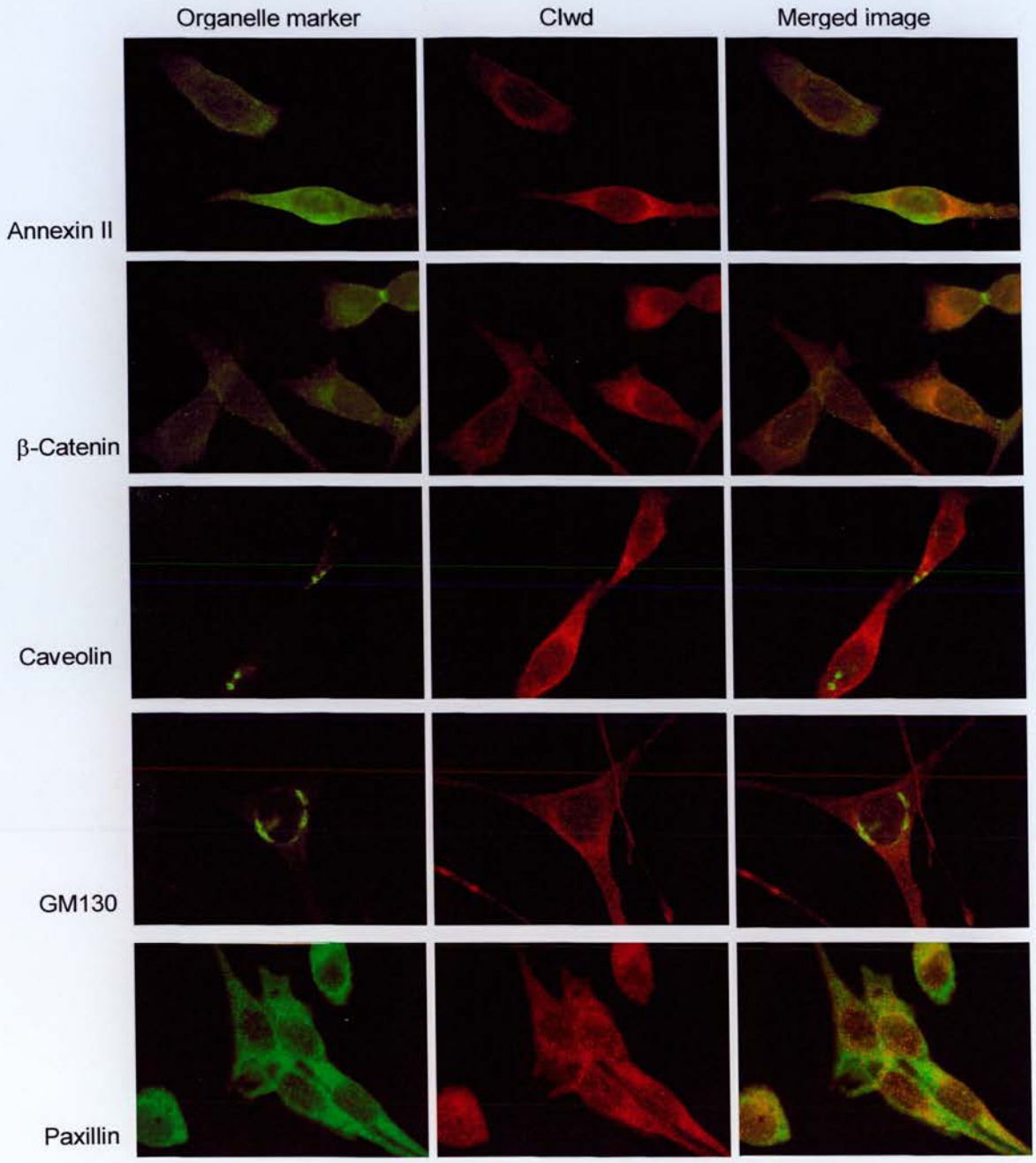


Figure 5.2 Double labelling of NIH/3T3 cells with Clwd antibody and various organelle markers

While the staining of GM130, Caveolin and Paxillin appear specific the cells stained with the anti-annexin II antibody and the β-catenin appear to have a far broader expression pattern than one would expect from antibodies staining the plasmalemma and zonula adherens. While the panels in figure 5.2 show cells stained for both the organelle marker and Clwd at the same time, cells stained for the organelle marker alone gave a different staining pattern, as can be seen in figure 5.3. It may be that there is some cross reaction between these particular antibodies and the Clwd antibody. Either way, neither expression pattern is similar to Clwd.

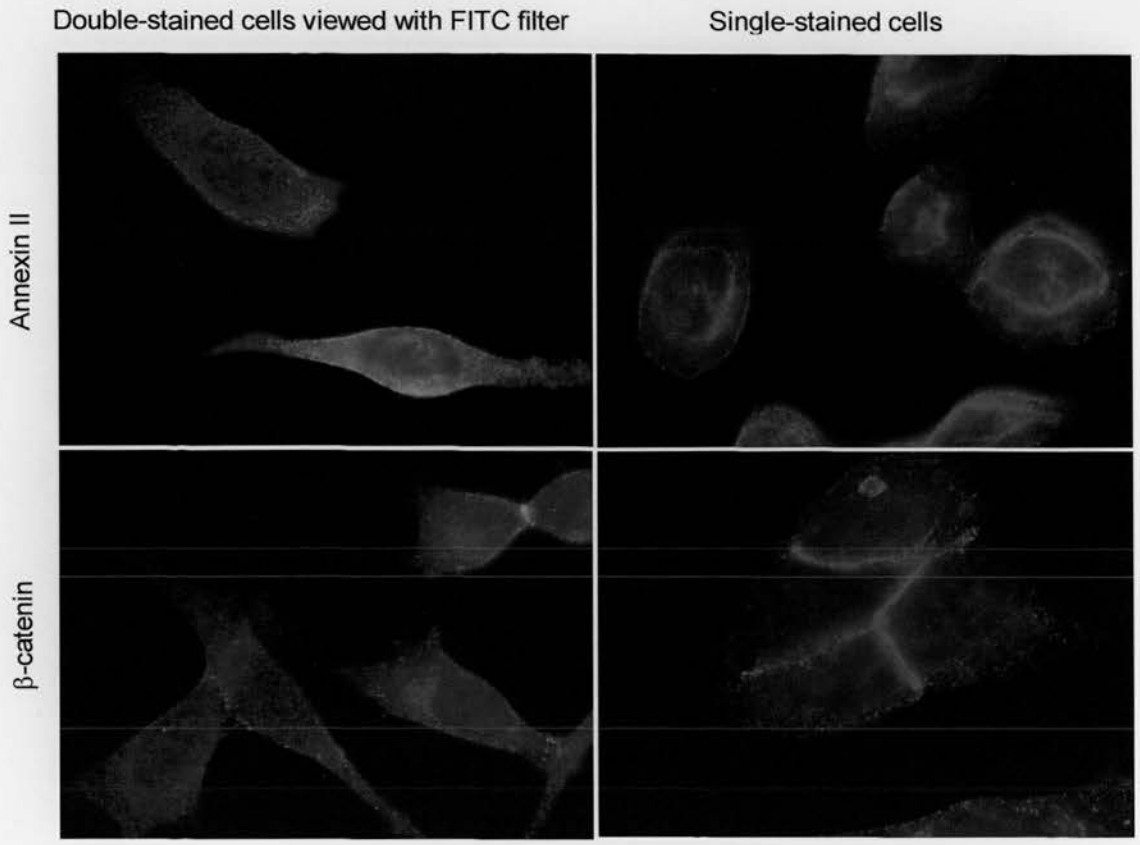


Figure 5.3 NIH/3T3 cells stained with anti-Annexin II antibody and anti-β-Catenin antibody alone or in conjunction with anti-Clwd antibody

The next figures are the results of co-localisation experiments firstly with additional Golgi apparatus markers (figures 5.4 and 5.5) and secondly with markers of the nuclear envelope such as Lamin A/C (figures 5.6 and 5.8) and Nuclear pore O-linked glycoprotein (figure 5.7). With Clwd expression mainly in the cytoplasm but concentrated at the nuclear envelope both the Golgi apparatus and the components of the nuclear envelopes were good candidates for the position of Clwd. These experiments were carried out in both HeLa and NIH/3T3 cell lines as noted in the figures.

The Golgi apparatus is a collection of flattened membranes, or cisternae, and vesicles. Proteins mature and are modified after their biosynthesis at the ER and transported through the Golgi apparatus. This involves the transport of protein-containing vesicles through a process of vesicle budding and fusion from the cis-compartment (CGN) to the medial-compartment (MGN) and the trans-compartment (TGN) of the Golgi apparatus.

While the Golgi marker GM130 did not co-localise well with Clwd (figure 5.2), given its location within the cis-Golgi compartment and the fact that the Golgi is a large organelle, I felt it was important to examine other components. In this experiment, I examined the localisation of Vti1a (figure 5.4) and Vti1b (figure 5.5). Both these proteins are involved in vesicle transport, with Vti1a functioning in protein transport within the secretory pathway and Vti1b involved in the regulation of post-Golgi vesicle trafficking. As these proteins are found in trafficking vesicles, it is possible that they have a more diffuse pattern of expression than GM130 and so would be better candidates to be co-localised with Clwd

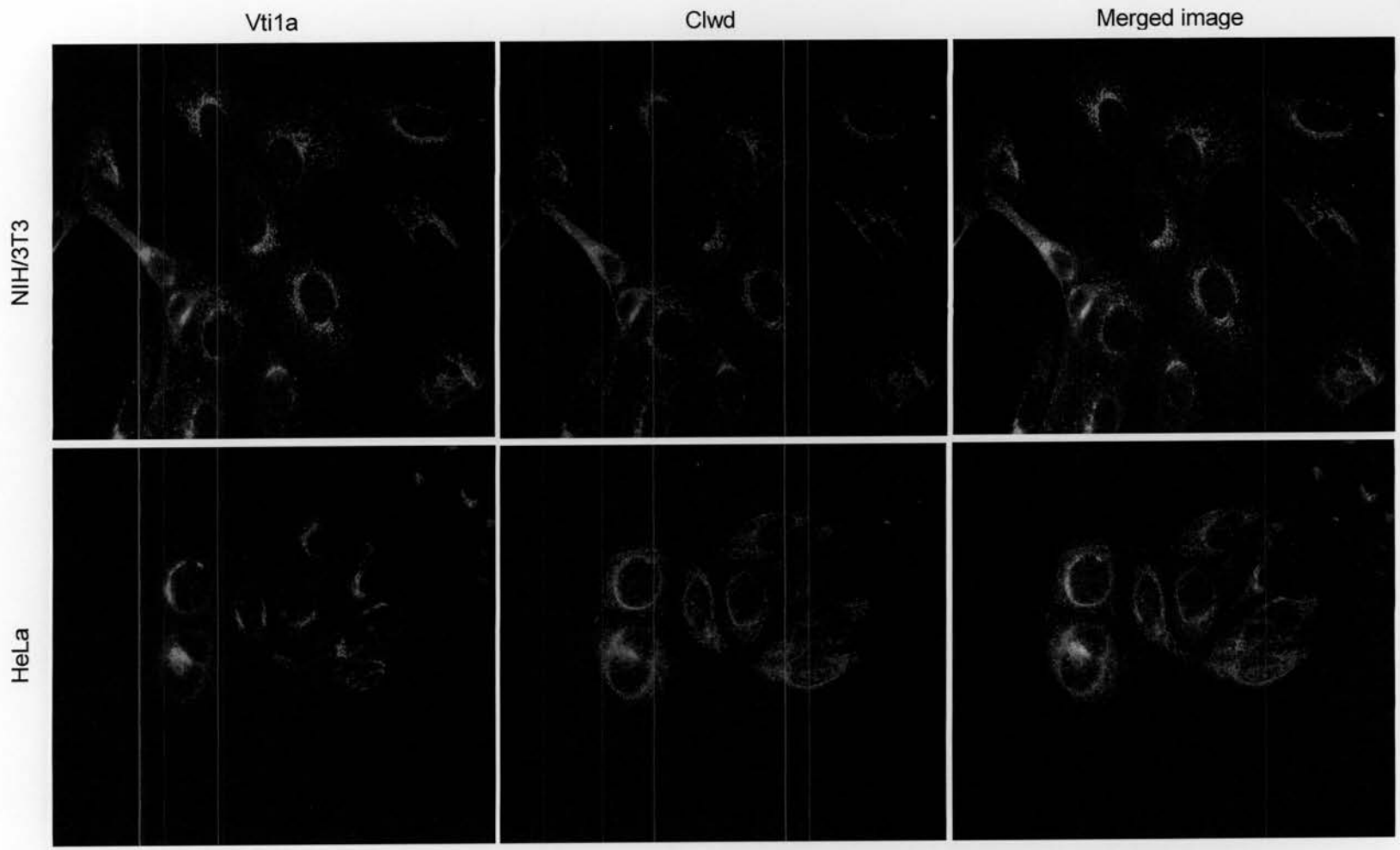


Figure 5.4 HeLa and NIH/3T3 cells stained with anti-Vti1a antibody.

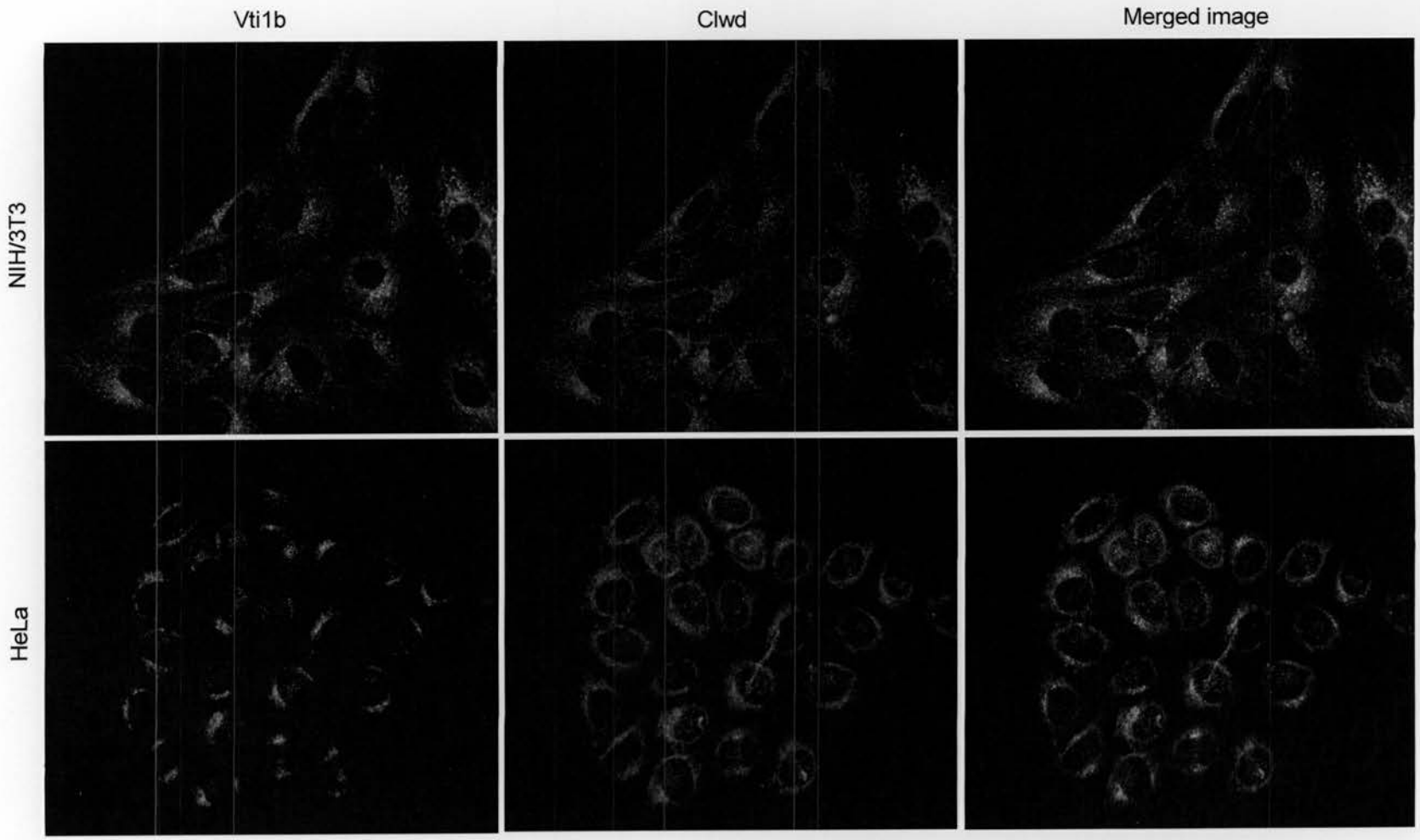
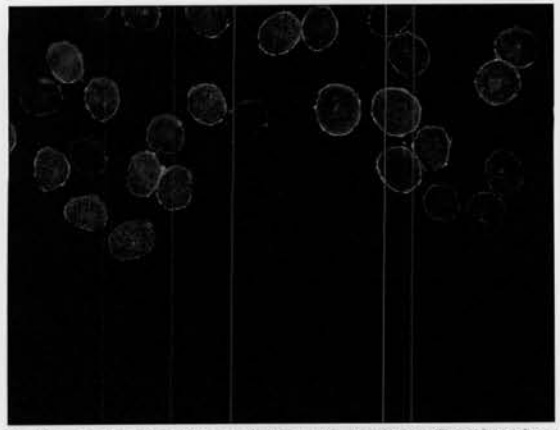


Figure 5.5 HeLa and NIH/3T3 cells stained with anti-Vti1b antibody.

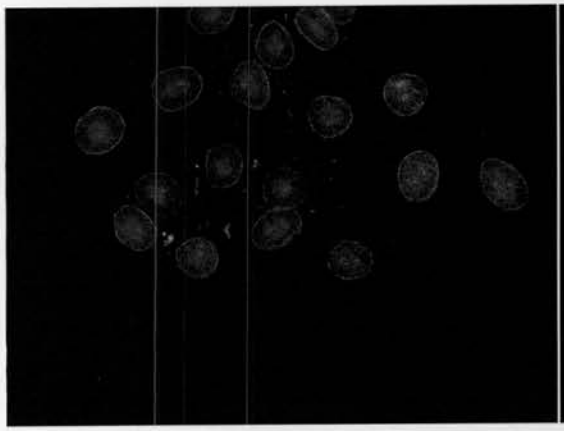
In NIH/3T3 cells Vti1a and particularly Vti1b expression is seen in the cytoplasm more than restricted to the cis-Golgi as GM130 was, and so is closer to the pattern of Clwd expression. However in HeLa cells Vti1a and b expression is more similar to that of GM130 and found to one side of the nucleus. Clwd expression, in contrast, surrounds the nucleus. It seems unlikely therefore that Clwd is part of the Golgi apparatus

Given the staining pattern of Clwd seen surrounding the nucleus, markers of the nuclear envelope were the next tested. The markers chosen here were Lamin A/C and Nuclear pore O-linked glycoprotein. Lamin A/C is a type V nuclear lamin. The lamins are components of the nuclear lamina, a fibrous layer on the nucleoplasmic side of the inner nuclear membrane which is suggested to provide a framework for the nuclear envelope and may interact with chromatin.
Nuclear pore O-linked glycoprotein antibody actually detects up to eight different proteins from the nuclear pore complex (NPC). The NPC spans the dual membrane of the nuclear envelope and acts as a gateway for macromolecular traffic between the cytoplasm and the nucleus.

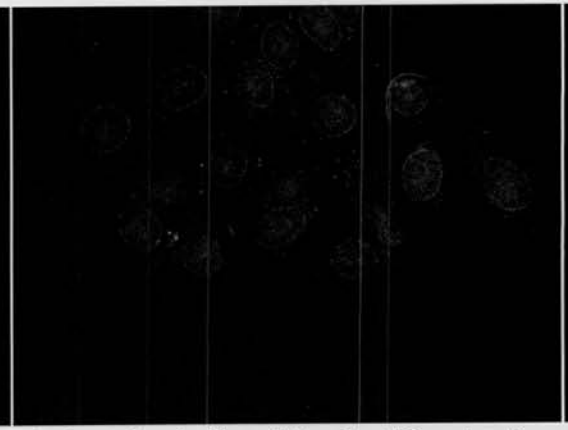
In these panels, Lamin A/C and Nuclear pore O-linked glycoprotein appear green, Clwd is red and where these co-localise is yellow.
The results of the double labelling can be seen in Figure 5.6, 5.7 and 5.8.



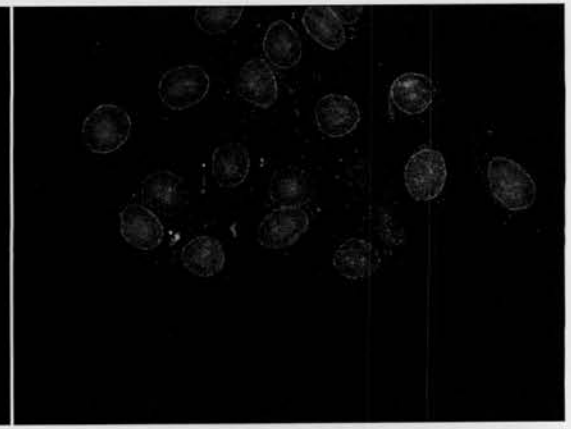
Cells stained with anti-Lamin A/C antibody alone



Cells stained with anti-Lamin A/C and anti-Clwd antibodies, FITC-filter only

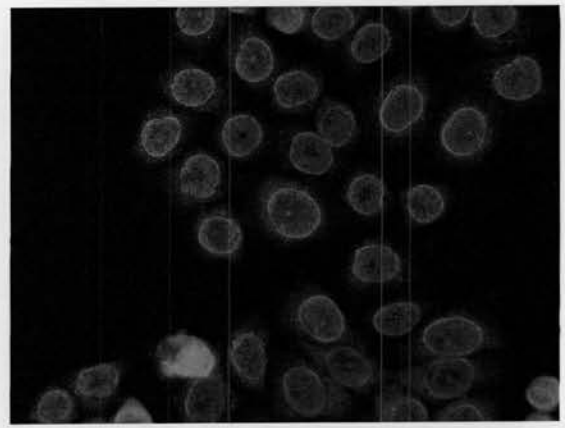


Cells stained with anti-Lamin A/C and anti-Clwd antibodies, Texas Red-filter only

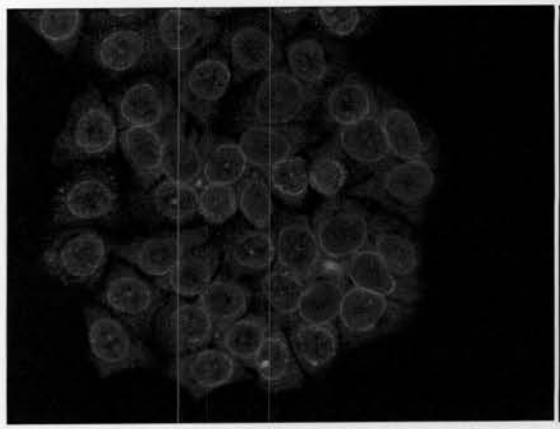


Cells stained with anti-Lamin A/C and anti-Clwd antibodies, merged image

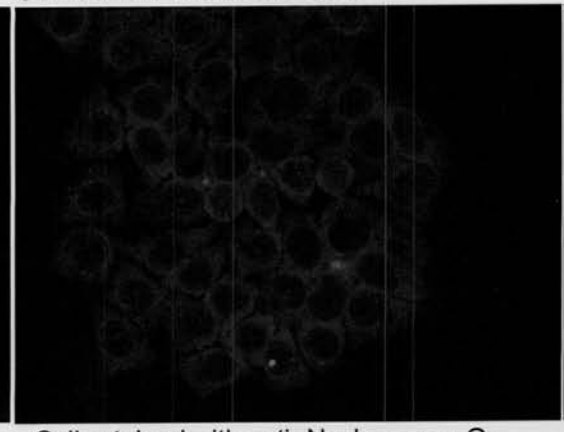
Figure 5.6 HeLa cells stained for Clwd and Lamin A/C.



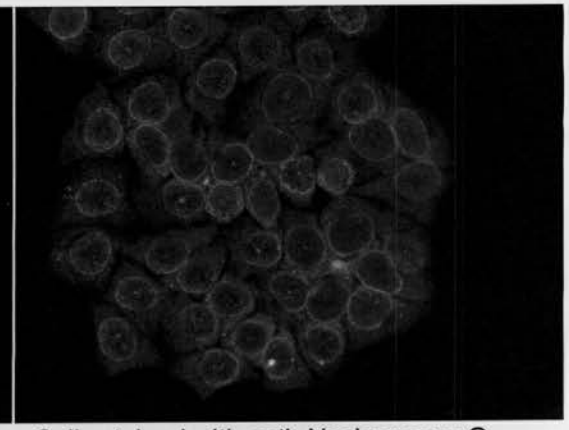
Cells stained with anti- Nuclear pore O-linked glycoprotein C antibody alone



Cells stained with anti- Nuclear pore O-linked glycoprotein and anti-Clwd antibodies, FITC-filter only



Cells stained with anti- Nuclear pore O-linked glycoprotein and anti-Clwd antibodies, Texas Red-filter only



Cells stained with anti- Nuclear pore O-linked glycoprotein and anti-Clwd antibodies, merged image

Figure 5.7 HeLa cells stained for Clwd and Nuclear pore O-linked glycoprotein.

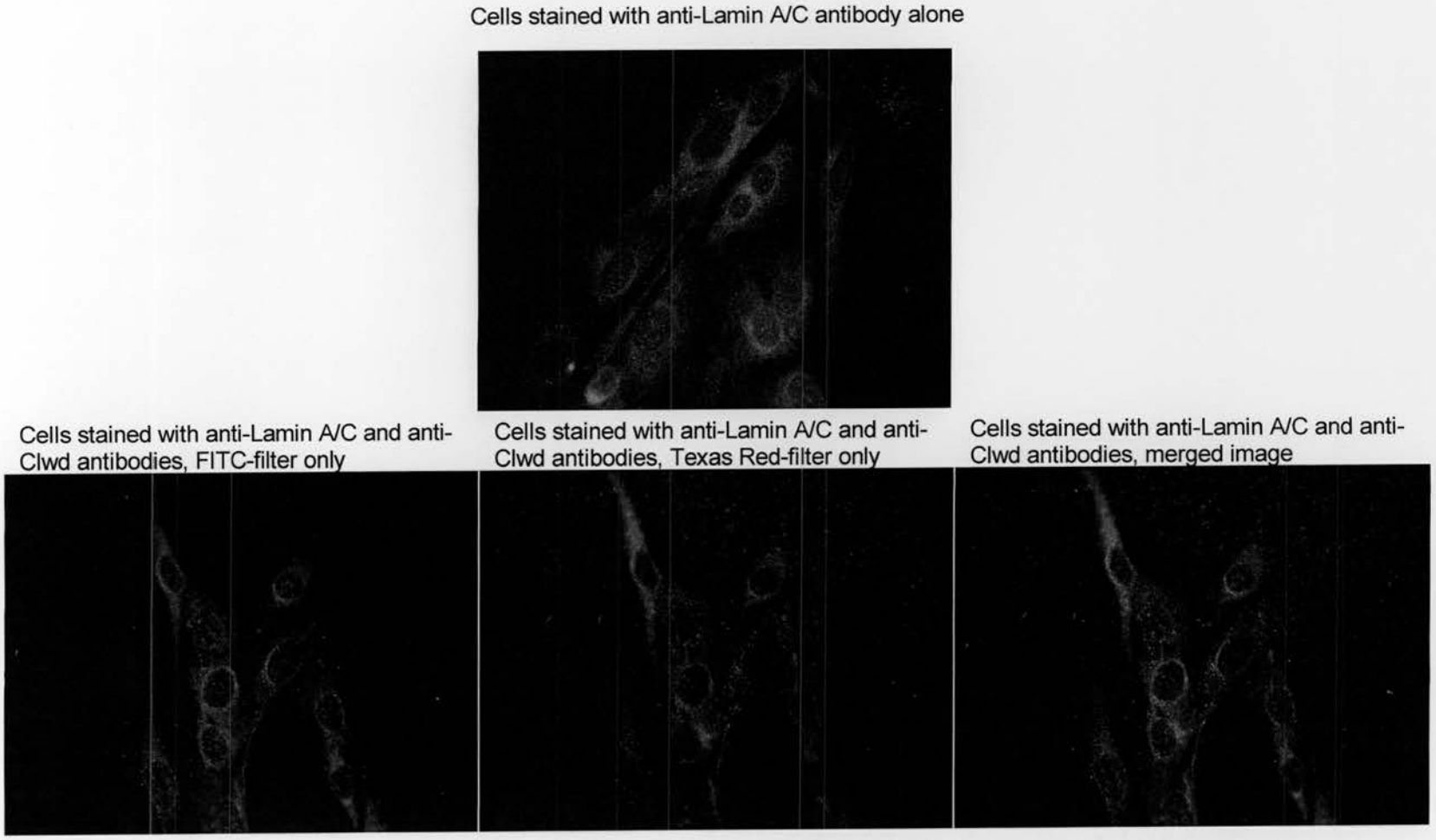


Figure 5.8 NIH/3T3 cells stained with Clwd and Lamin A/C.

In this co-localisation experiment, in both cell lines and antibodies, there appears to be some cross-reaction between the antibodies. The expression pattern of Clwd in the HeLa cells stained with anti-Clwd and anti-Lamin A/C antibodies (figure 5.6) is completely different from that seen previously as there is Clwd-staining in the nucleus. In addition the Lamin A/C staining differs depending on whether or not Clwd is present. In the NIH/3T3 experiment (figure 5.8) however, LaminA/C appears to have a completely different expression pattern, one very similar to that of Clwd. As the expression pattern in the cells stained with anti-Lamin A/C alone is similar to that in the double labelling experiment, it does not appear that the cross-reaction seen in HeLa cells is relevant here. Why this should be so is unclear.

In the nuclear pore co-localisation (figure 5.7), the expression pattern of the proteins recognised by the anti-Nuclear pore O-linked glycoprotein antibody is again quite similar to Clwd although it is possible, in the merged image panels, to see that Clwd is found in the cytoplasm around the borders of the nucleus whereas the nuclear pore complex proteins are not.

5.3 Gene silencing by RNAi

5.3.1 Design of siRNA

Strict rules need to be applied to the design of the siRNAs to ensure that they are both efficient and specific for the gene in question. The rules I adopted were those described by the Tuschl lab at their website "The siRNA User Guide" which can be found at http://www.rockefeller.edu/labheads/tuschl/sirna.html . These state that the target sequence of the gene in question should begin 50 to 100 nucleotides downstream of the start codon, and the sequence used should have the structure AA(N19)TT where N is any nucleotide. The target sequence should have approximately 50% G/C content and this sequence should be unique to the gene of interest to eliminate cross-reactivity.

The siRNAs chosen for Clwd were Clwd^{+89} `aagtgccgagtaccctgggga` and Clwd^{+183} `aaatccacactcccgttcatg` (see figure 5.9) which were 89 and 183 nucleotides from the start codon respectively. While there were no suitable AA(N19)TT sites, I used the next option recommended AA(N21). Clwd^{+89} had 68% GC content and Clwd^{+183}, 52% GC content. Using BLAST to compare these to other mouse cDNA sequences it was found that both were specific for Clwd. Having designed these target sequences the sense and antisense siRNA sequences were ordered from Xeragon along with a negative control siRNA available from the Xeragon library.

Clwd^{+89}:

Target sequence 5'-aagtgccgagtaccctgggga-3'

Sense siRNA: 5'-r(gugccgaguacccugggga)dTT-3'

Antisense siRNA 5'-r(uccccaggguacacggcac)dTT-3'

Clwd^{+183}:

Target sequence: 5'-aaatccacactcccgttcatg-3'

Sense siRNA: 5'-r(auccacacucccguucaug)dTT-3'

Antisense siRNA 5'-r(caugaacgggaguguggau)dTT-3'

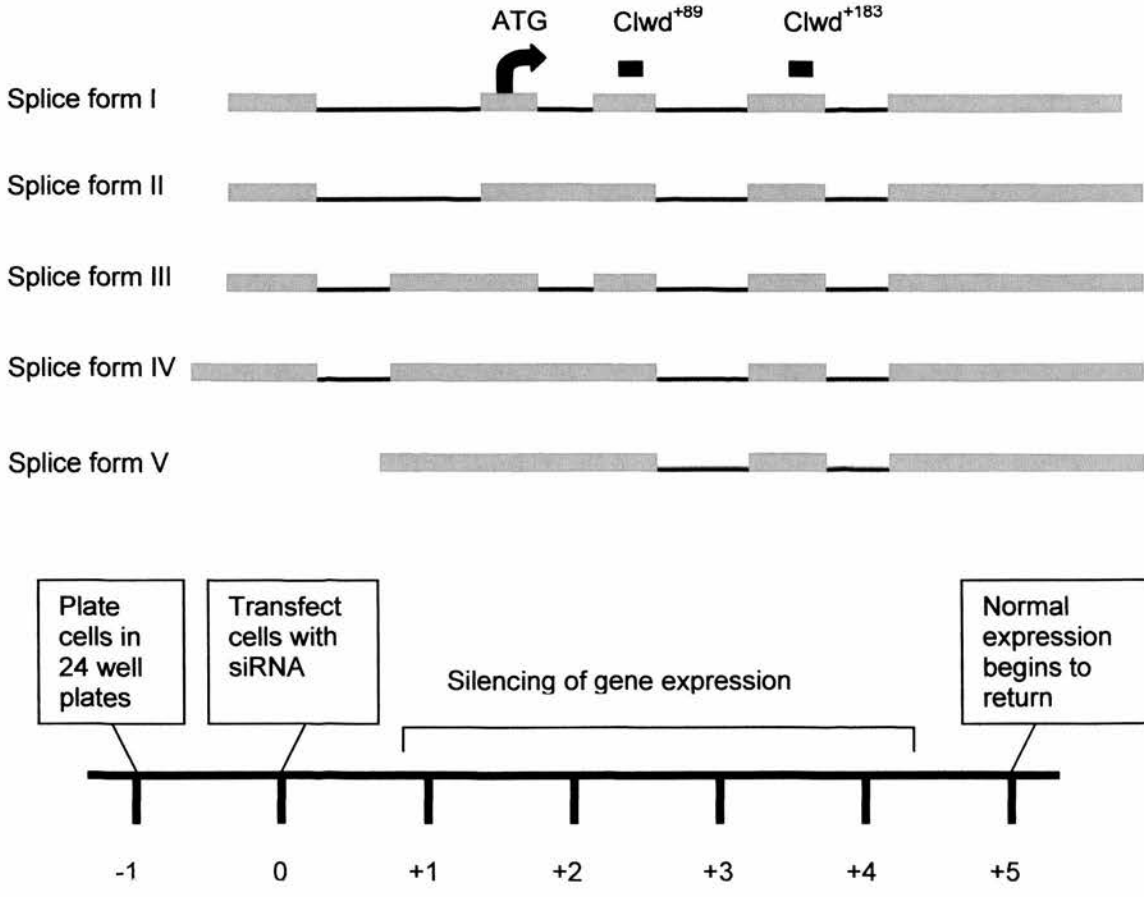


Figure 5.9 Diagrams showing the siRNAs used to silence Clwd, and the timeline of an RNAi experiment.

5.3.2 Knock-down of Clwd expression using RNAi

Upon receipt of the siRNAs, the sense and antisense strands were annealed and transfected into NIH/3T3 cells as described in chapter 2. The cells were plated on day -1 in 24-well plates and transfected on day 0 with samples taken from day 2 to day 5 as shown in figure 5.9. The NIH/3T3 samples taken on these days belonged to 5 different treatment groups as described below:

No treatment:

Cells were not transfected at all and once plated remained untouched until samples were taken on the indicated days.

Mock transfection:

These were cells that were subjected to Oligofectamine™ and OptiMEM® as for the transfected cells, except with no siRNA present. This control served to judge the effect the transfection procedure had on the cells.

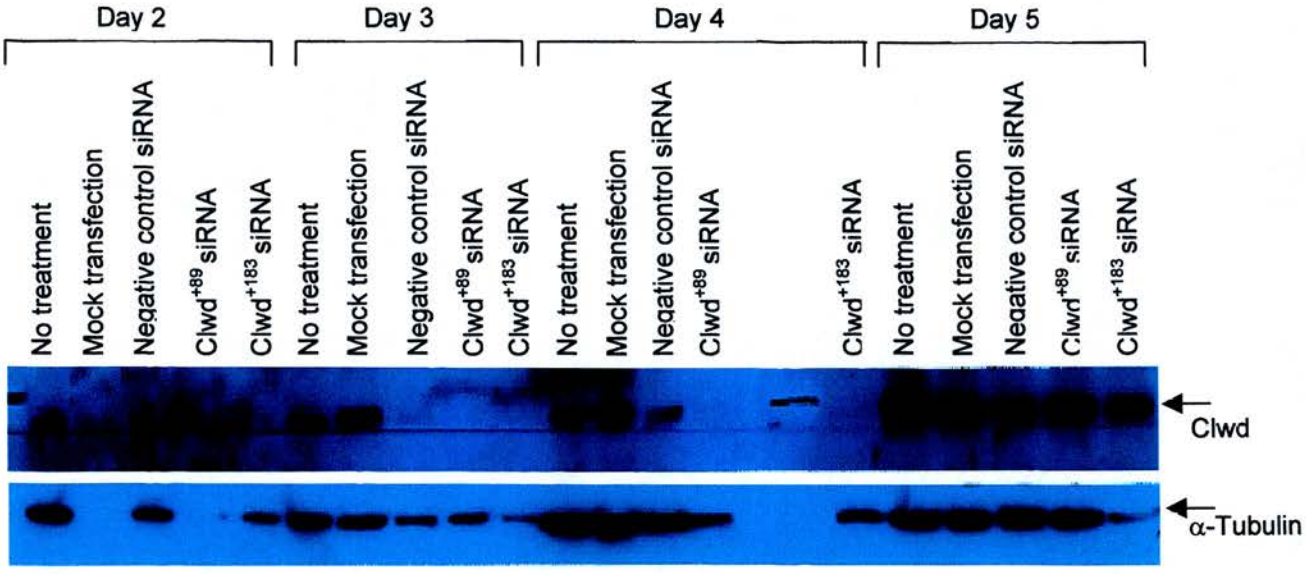
Control siRNA:

These cells were transfected with the control siRNA. This would show the effect of random siRNA molecules, with no specificity for any gene, on the cell.

$Clwd^{+89}$ siRNA/$Clwd^{+183}$ siRNA:

This was the transfection of the siRNAs specific for Clwd as described earlier. Any phenotype seen in these cells and not in the control cells would be deemed a result of the knock-down of Clwd expression.

In the first RNAi experiment the efficiency of the Clwd siRNAs was tested. This was done by extracting the protein from the transfected cells, 2-5 days post-transfection and using the anti-Clwd antibody to assess Clwd expression on a Western blot of these extracts. α–Tubulin was the loading control for the Western blot, and (as for the RT-PCR experiment in chapter 4) Scion Image software expressed the level of Clwd expression in relation to the level of α–Tubulin expression. The results of this experiment are shown in figure 5.10.



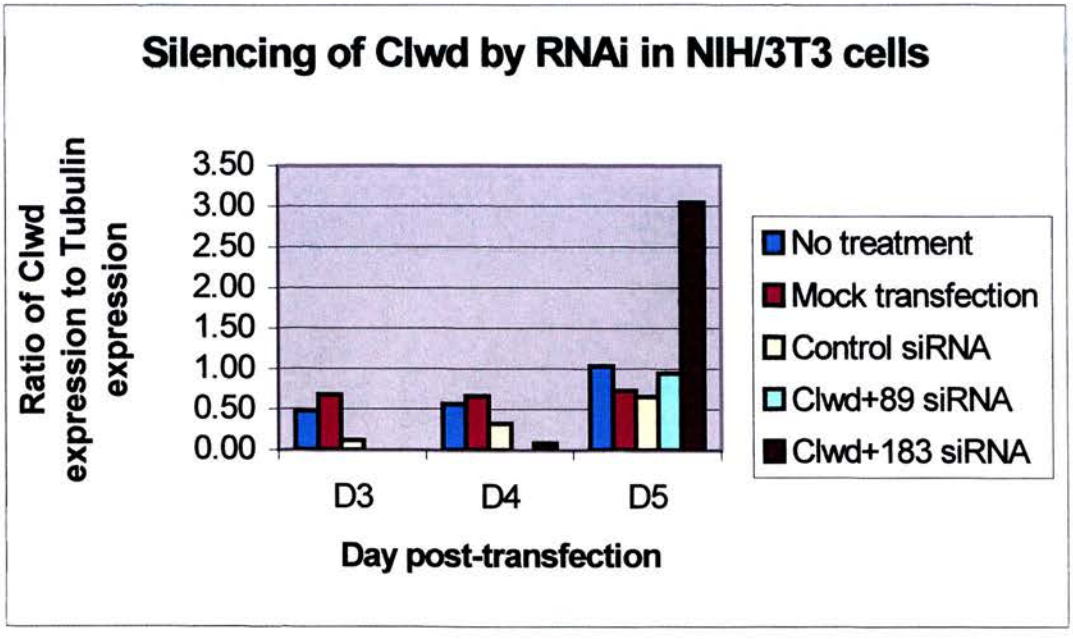


Figure 5.10 Western blot of the protein extracted from cells transfected with siRNA and histogram describing Clwd expression in these extracts. In this experiment, Clwd expression in reduced in the Clwd^{+89} and Clwd^{+183} samples on days 3 and 4 in comparison to the negative controls while on day 5 the expression of Clwd appears to have recovered.

The graph in figure 5.10 shows the relative expression of Clwd to α–Tubulin on sample days 3 to 5 (given that the expression of both Clwd and α–Tubulin is extremely low on day 2 this is not shown in the graph). The first thing to notice is that it appears that Clwd expression is reduced in the $Clwd^{+89}$ and $Clwd^{+183}$ samples on days 3 and 4 in comparison to the negative controls, while on day 5 it appears that expression has recovered. However it is worth noting that it appears that Clwd expression is reduced in the control siRNA-treated samples also. Although this reduction is never as great as that in the Clwd siRNA-treated samples it is important as the sequence of the control bears no relation to Clwd (or any other mouse RNA sequence) and so how it can apparently reduce Clwd expression is unclear. One explanation could be that the transfection of siRNA is having some broad general effect on the state of the cell and as a result is having a secondary effect on Clwd. This control siRNA has been used by other groups successfully however and so does not seem to be a general feature of this particular sequence of siRNA. Nevertheless in my hands the transfection of siRNA appears to have some effect on the cells in addition to the effect of a mock transfection.

Despite this anomaly there is a further reduction of Clwd expression in both Clwd siRNA-treated samples, and the fact that the transfection of two different siRNA sequences targeting the same gene result in a decrease in Clwd expression in comparison to the control siRNA is highly significant in attempting to assess the validity of the RNAi experiment.

Following this result, it was important to determine the phenotype associated with this reduction in Clwd expression. To accomplish this, the cells were plated on coverslips in 24-well plates in order that the cells could be stained by the anti-Clwd antibody and visualise the Clwd-silenced cells by immunofluorescence. The results of this experiment can be seen in figure 5.11.

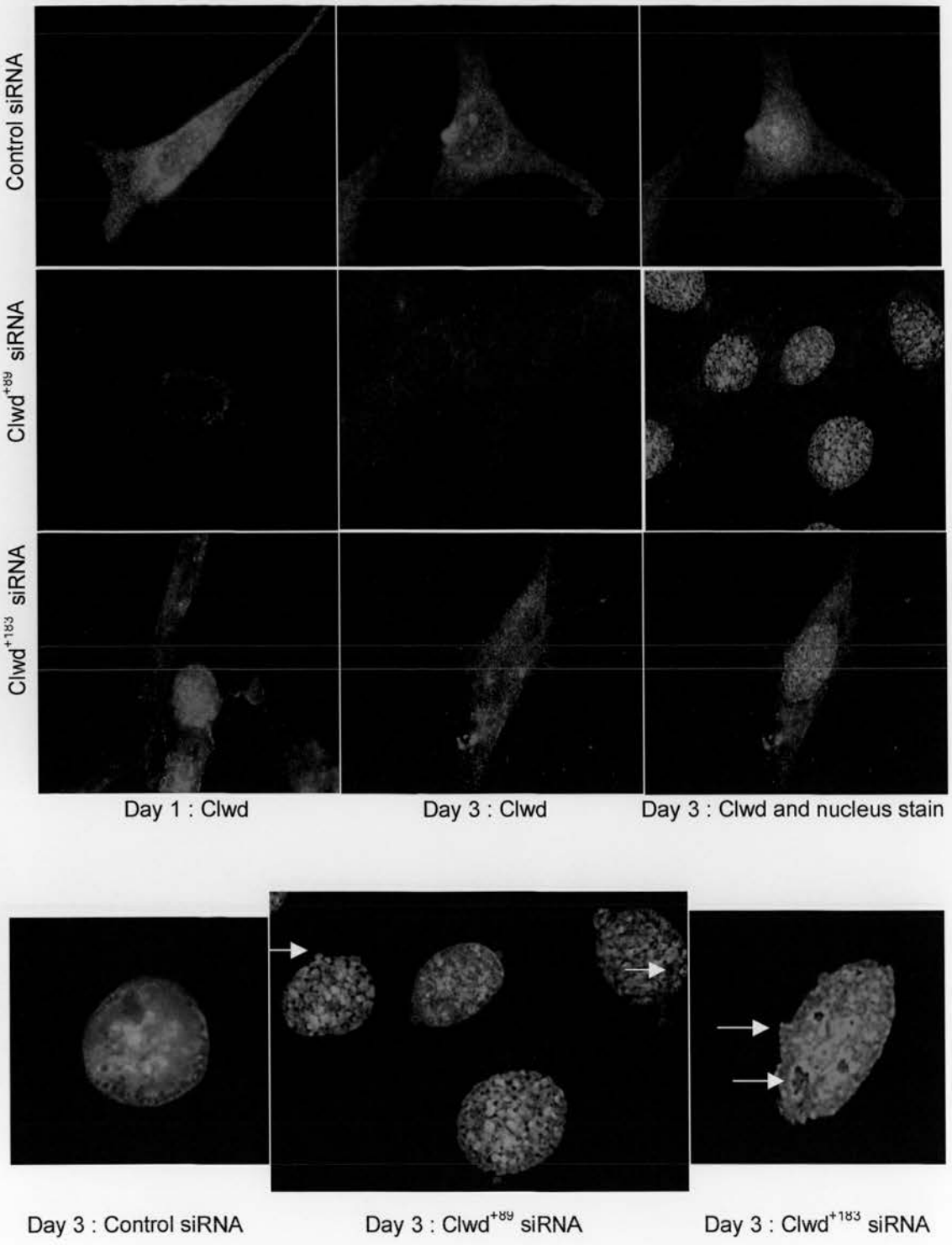


Figure 5.11 Clwd silencing in NIH/3T3 cells. Lower panels compare nuclei from negative control siRNA transfected cells and anti-Clwd siRNA transfected cells. Arrows highlight apparent vacuoles and protrusions.

The advantage to viewing the immunofluorescence image of an RNAi experiment is that it is possible to examine on a cell by cell basis if the gene expression has been reduced and the effect, if any, on that cell. This is an improvement on determining the overall expression on a population of cells as not all cells are necessarily transfected and even a cursory evaluation of phenotype is impossible.

As can be seen from the immunofluorescence results seen in figure 5.11, Clwd^{+89} decreases Clwd expression in the cell. It should be noted that all of the images above were taken with the same fixed exposure so that the intensity of the staining could be directly compared. The amount of Clwd expression did not seem to be as decreased in the Clwd^{+183}. There was a phenotype associated with the loss of Clwd. There were far fewer cells in the wells treated with both the Clwd siRNAs than in the controls. The nuclei of the Clwd siRNA-treated cells were irregular in shape with protrusions and vacuole-like gaps in DAPI staining – this was not seen in the control siRNA-treated cells. These aspects of the Clwd siRNA treated cells were seen as early as one day post-transfection and were still visible at 2 days later.
It seemed therefore that the loss of Clwd was having an effect upon either the growth or survival of the cells and upon the structure of the nucleus

5.3.3 Problems with RNAi

Unfortunately, in the remaining time since these experiments were carried out, I have been unable to reproduce Clwd silencing using siRNA. I addressed the possible reasons for this in a variety of ways.

In an attempt to address the potential problem of transfection efficiency, NIH/3T3 cells were ordered from ECACC. These would be of low passage number and guaranteed mycoplasma-free, two important factors in maintaining efficient and consistent transfection. I maintained this cell line in the recommended medium without antibiotics and without allowing it to reach confluency. In addition, a positive control siRNA was ordered from Qiagen that had been proven to be efficient in silencing Vinculin expression in NIH/3T3 cells. Using these fresh, low passage number cells and the positive control siRNA I tested a variety of siRNA transfection reagents described in RNAi experiments and the effect of the initial cell density on the success of the experiment. The results can be seen in figure 5.12. It can be seen that the best transfection reagent appears to be RNAiFect^{TM} (Qiagen). This delivered the siRNA which reduced Vinculin expression by half in all bar one of the days that samples were taken. Using this agent and the recommended protocol from the manufacturer, an RNAi experiment to silence Vinculin and Clwd was carried out and the cells analysed by immunofluorescence, but unfortunately neither siRNA achieved silencing. Whether this was an intermittent transfection problem, or the siRNAs being used were contaminated or had deteriorated in some way is unknown, and unfortunately time did not permit me to carry out further experiments.

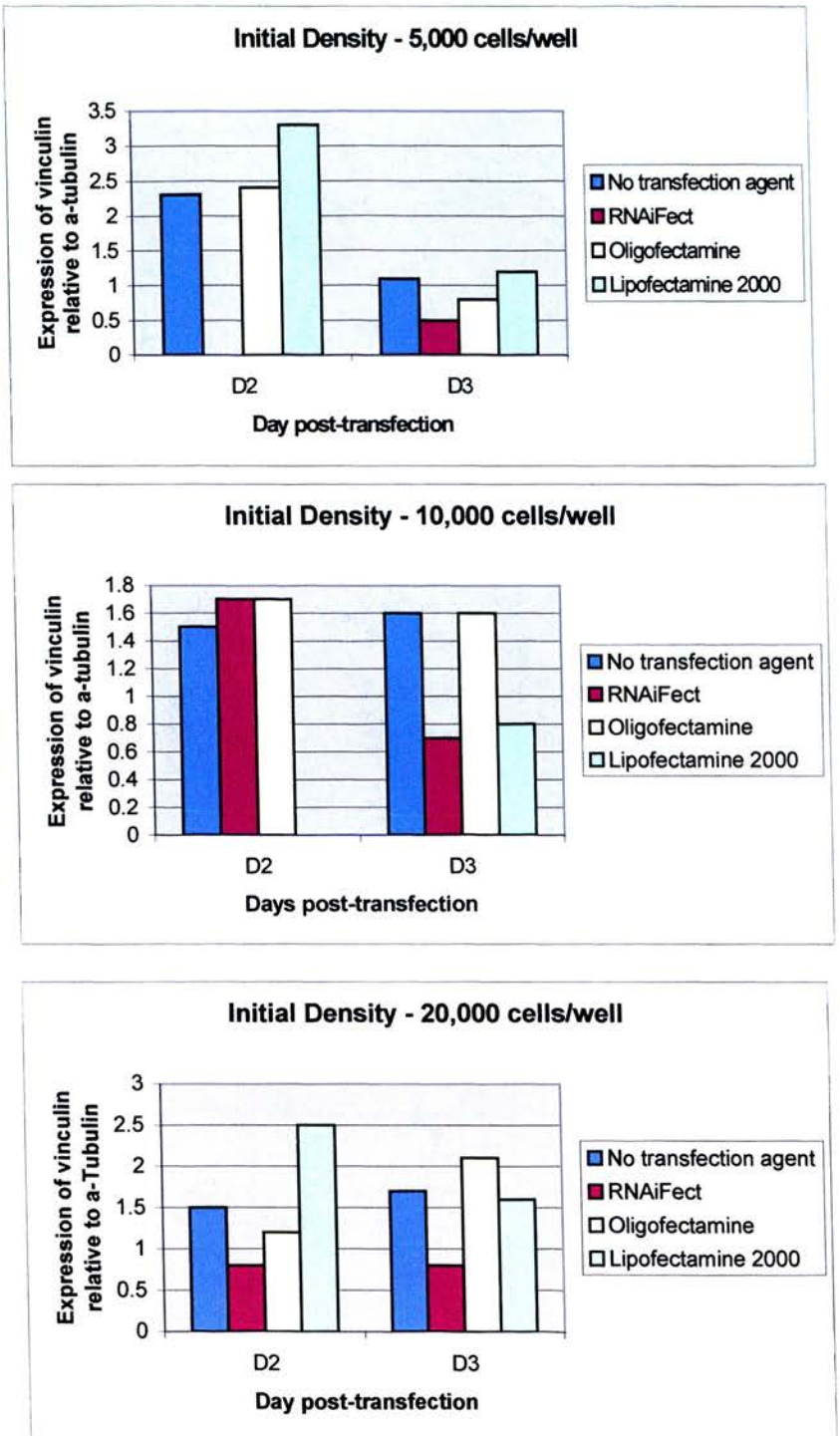


Figure 5.12 Efficiency of transfection reagents in delivering Vinculin siRNA – RNAi positive control.

5.4 Discussion

In this chapter I have described the steps I took to gain more information about the role Clwd might play in the cell. To do this I used two techniques, immunofluorescence to determine subcellular localisation and gene silencing in cell lines by RNAi

5.4.1 Subcellular Localisation

Using the anti-Clwd antibody described in chapter 4 and a secondary antibody with a fluorescent tag, it was possible to visualise Clwd in cell lines and make an assessment as to its location in the cell. In both NIH/3T3 cells and HeLa cells Clwd appears to be located in the cytoplasm with more concentrated staining surrounding the nucleus. Using an organelle marker system of antibodies Clwd appeared to have an expression pattern in the cells most similar to that of Paxillin.

Paxillin is a multi-domain protein that localises primarily to sites of cell adhesion to the extracellular matrix called focal adhesions in cultured cells. These focal adhesions form a structural link between the extracellular matrix and the actin cytoskeleton and are also important sites of signal transduction as Paxillin, like other focal adhesion proteins, serve as a point of convergence for signals resulting from stimulation of various classes of growth factor receptor.

This co-localisation of Clwd with Paxillin could suggest therefore that Clwd is either one of the other structural proteins, similar to Paxillin at the focal adhesions of the cell, or is it one of the signalling molecules taking part in a pathway situated at the site of Paxillin.

Given Clwd's small size, potential for phosphorylation/O-GlcNAcylation and lack of any known structural domains, I would suggest it is more likely to be a small undescribed signalling molecule, perhaps activated and inactivated by phosphorylation rather than a structural protein.

The co-localisation experiments with the Golgi markers and nuclear envelope markers were not as clear-cut. In some examples it appeared that each antibody,

when used to stain cells simultaneously, influenced the staining of the other. The reason for this is unclear, other than there is a form cross-reaction taking place with the secondary antibodies recognising both primaries even though they have supposedly different specificities.

5.4.2 RNAi

In this chapter I also described the use of RNAi in order to assess the effect on a cell once Clwd expression has been silenced. A decrease in Clwd expression was detected by Western blot and upon staining the cells by immunofluorescence. This confirmed that the siRNA designed had worked well. It also provided further evidence that the anti-Clwd antibody produced was specific for Clwd as transfecting Clwd-specific siRNA molecules resulted in a decreased signal from the antibody. The immunofluorescence experiment allowed an initial assessment of the cell phenotype resulting from reduced Clwd expression. Upon examination of the Clwd-silenced cells, there were far fewer cells than in the negative control siRNA-transfected cells. This could imply that there was a decrease in growth or survival of the cells. In addition the morphology of the DAPI-stained nucleus appeared unusual in those cells transfected by the anti-Clwd siRNA. There appeared to be gaps in the staining such that the nuclei seemed to exhibit vacuoles. There were also protrusions of DAPI-stain extending from the usually tightly formed oval structure of the nucleus as seen in the control siRNA-transfected cells

CHAPTER 6

IMMUNE SYSTEM PHENOTYPE OF WASTED MICE

6. Immune System Phenotype of Wasted Mice

6.1 Introduction

In this chapter I will describe a preliminary investigation into the cause of the immune system defects seen in wasted mice as described in chapter 1. Clwd was examined initially as it was situated close to the wasted deletion and so could play a role in the pathology of wasted mice. In particular, as there appeared to be no role for eEF1A2 in the immune system, it was thought that the abnormalities here could be attributable to Clwd.

I will introduce the results of experiments examining this hypothesis with the work of Dawn Loh who created the transgenic animals I used to assess the role of Clwd in the wasted mice. This transgenic experiment was designed in the first instance to rescue wasted mice by supplying an intact copy of the genomic region surrounding the wasted deletion. The construct used was a genomic BAC clone (RPCI-22-219F24 referred to as 219F24) that carried: the first exon of a voltage-gated potassium channel gene *Kcnq2, Eef1a2, Clwd* and the protein tyrosine kinase gene *Ptk6* but without its 5' UTR and promoter sequences. It was hoped that this would replace the lost functionality of the wasted deletion and thus rescue the wasted mice. To test what portion of the wasted pathology was attributable to the loss of *Eef1a2*, the same genomic BAC clone was engineered to have a deletion within *Eef1a2* removing exons 2-4 and thus the start codon for the protein and all three GTP-binding domains essential for eEF1A2 function (referred to as del219F24). If eEF1A2 was solely responsible for the wasted pathology then while the intact BAC, 219F24, would rescue the wasted mice, the BAC with a portion of *Eef1a2* deleted, del219F24, would not. However if Clwd were partially responsible for the wasted phenotype then wasted mice carrying del219F24 would have a partially corrected phenotype *e.g.* if Clwd were responsible for the immune system abnormalities, it could be expected that these mice would still develop muscle wasting and tremors, but their spleen and thymus would be as normal. See figure 6.1 for a diagram of these constructs.

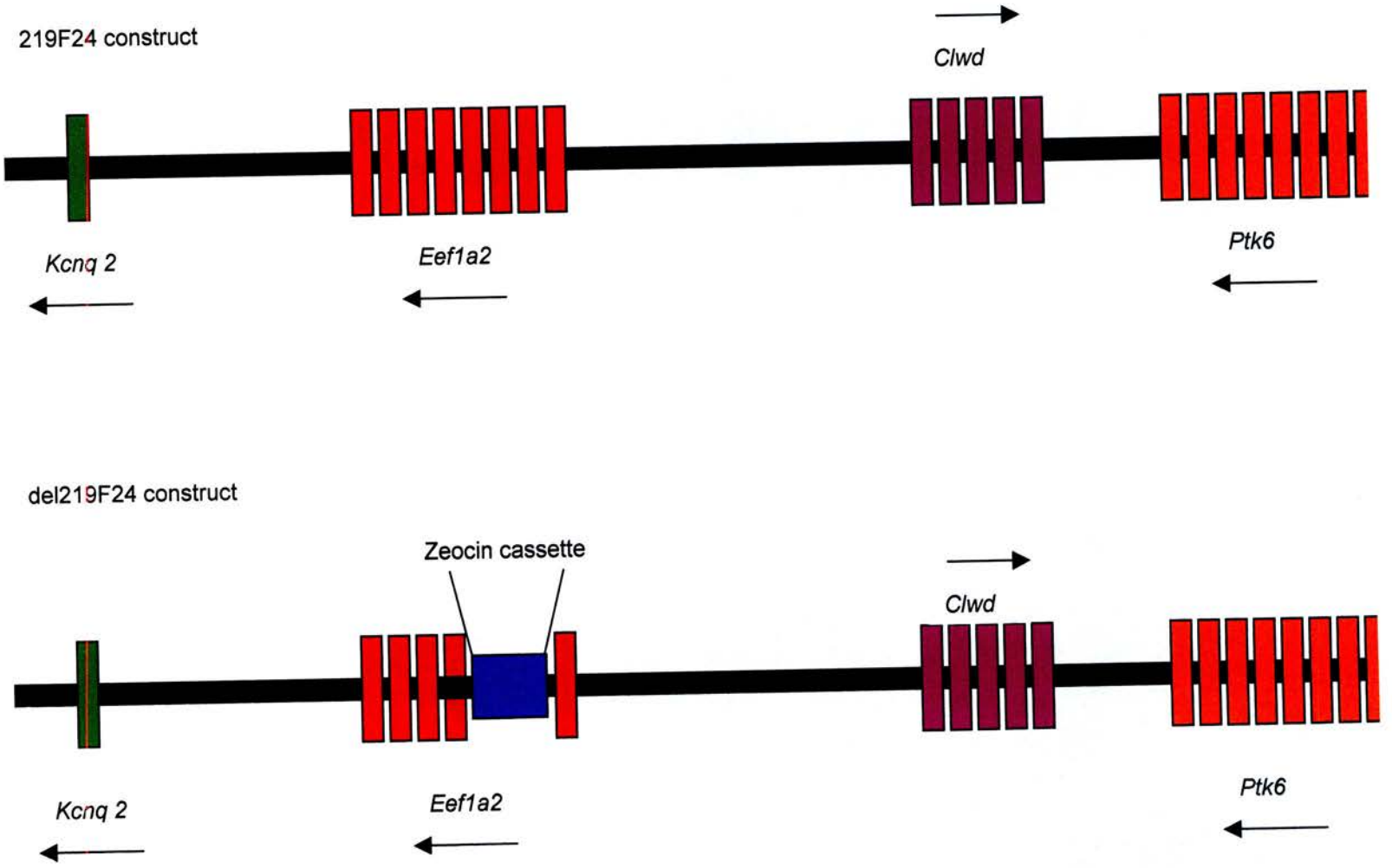


Figure 6.1 Diagram of transgenic constructs

In order to genotype animals carrying these transgenes a system of PCR primers was used that could distinguish the wild type allele (P2) at the wasted locus, the wasted allele (wstspan) and the presence of the BAC constructs (B1162-3).
While the wild type allele primers were designed to amplify a portion of the wasted deletion sequence, the wasted allele primers consisted of a split forward primer (wstspan-f) and a reverse primer (wstspan-r). The split primer was designed with 18 nucleotides from just outside the 5'end of the wasted deletion and six nucleotides from just outside the 3' end of the wasted deletion. The regular primer was designed from exon 2 of *Eef1a2*. A positive result for this PCR would mean that the split primer would no longer be split due to the loss of the intervening wasted deletion sequence and so would imply the presence of the wasted allele. (See figure 6.2 below)

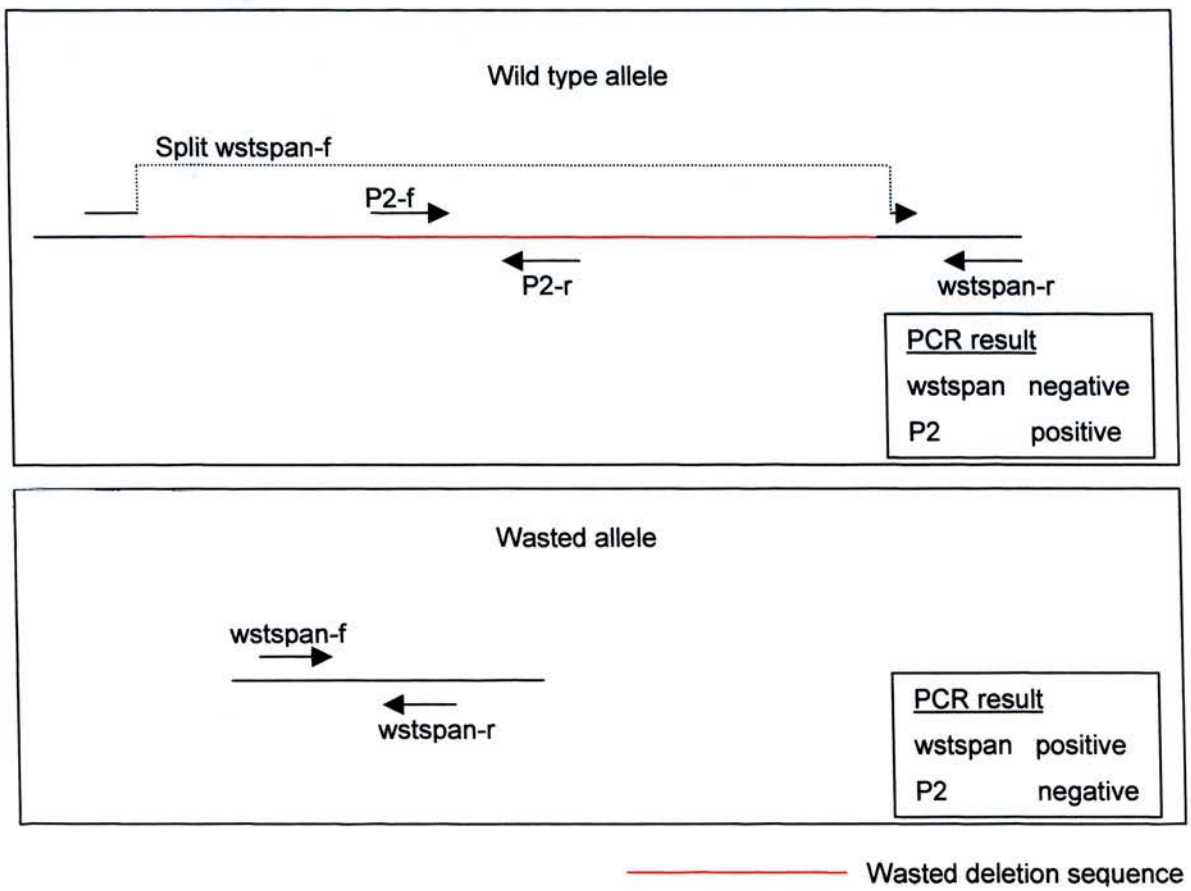
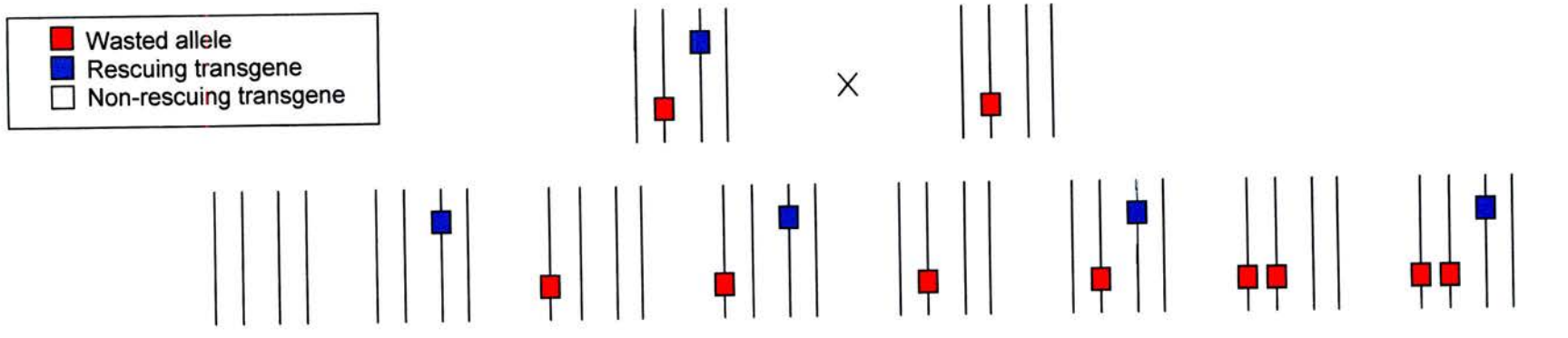


Figure 6.2 Diagram showing the primers for the genotyping PCRs

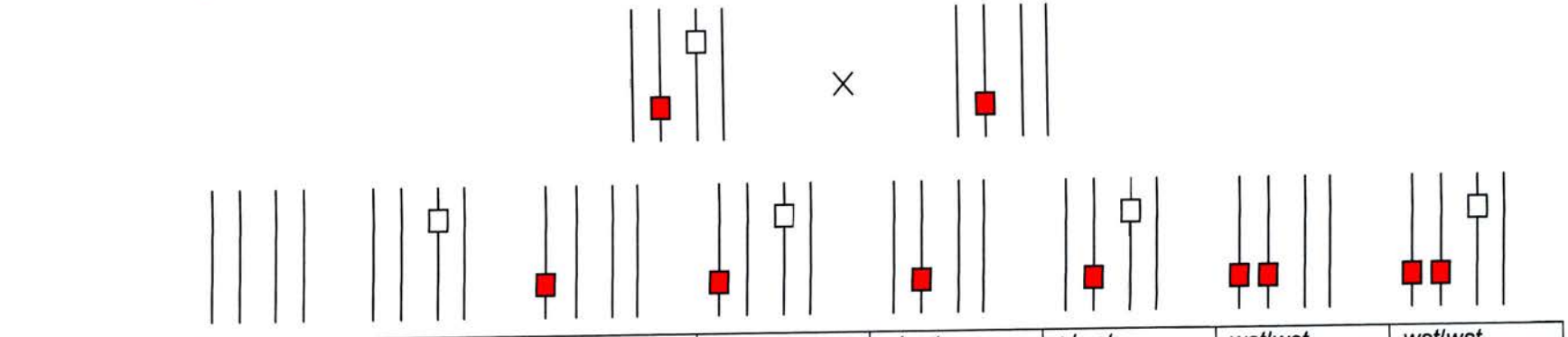
However this presented a problem as a genotyping strategy. As 219F24 was a genomic clone, it also represented a wild type allele as denoted by a positive P2 PCR result. Therefore, by PCR genotyping alone it could not be determined if a particular mouse positive for all three PCRs was an unaffected heterozygote carrying a transgene or a wasted homozygote rescued by the carried transgene. This is best explained in the table below.

wstspan	P2	B1162-3	Genotype
-	+	-	Wild type, non-transgenic
-	+	+	Wild type transgenic
+	+	-	Wasted heterozygote, non-transgenic
+	-	-	Wasted homozygote, non-transgenic
+	+	+	Wasted heterozygote, transgenic OR Wasted homozygote, transgenic

This meant that whether 219F24 and/or del219F24 rescued the wasted mice or not had to be determined by counting the number of wasted mice that were produced from crosses of wasted heterozygote mice, with one parent carrying the transgene. The normal frequency of a wasted mouse produced from a heterozygote cross would be one in four. This frequency would remain the same if the transgene did not rescue the wasted phenotype. However, if the transgene did rescue the phenotype, the number of wasted mice produced should be reduced to at least one in eight (given a copy number of one). The frequency of wasted homozygotes produced will remain at 1 in 4 but at least half of these homozygotes will carry a transgene and so while homozygous, will not result in a wasted mouse (see figure 6.3)



Genotype	+/+	+/+	+/*wst*	+/*wst*	+/*wst*	+/*wst*	*wst*/*wst*	*wst*/*wst*
Transgene?	-	+	-	+	-	+	-	+
Phenotype	Normal	Normal	Normal	Normal	Normal	Normal	Wasted	Rescued



Genotype	+/+	+/+	+/*wst*	+/*wst*	+/*wst*	+/*wst*	*wst*/*wst*	*wst*/*wst*
Transgene?	-	+	-	+	-	+	-	+
Phenotype	Normal	Normal	Normal	Normal	Normal	Normal	Wasted	Wasted

Figure 6.3 Diagram showing possible outcomes from transgenic, wasted heterozygote matings.

6.2 Is Clwd a good candidate?

6.2.1 Transgenic experiment to correct wasted phenotype

There were 4 transmitting founder mice for the 219F24 (A261, A262, A262.1, A262.2) and 1 transmitting founder for the del219F24 (A240). I established a breeding programme for each of these lines such that each line had at least one breeding pair with one parent carrying the transgene, and both heterozygous for the wasted allele.

The offspring from these matings were genotyped and monitored for the number of phenotypically wasted mice. Using binomial distribution it was possible to calculate the probability that the observed number of wasted mice produced showed the possible frequencies of 1 in 4 (0.25) or 1 in 8 (0.125). These data can be seen below.

del219F24

Line	Wasted Offspring	Total Offspring	P value at 0.25	P value at 0.125
A240	19	88	0.27	0.01

219F24

Line	Wasted Offspring	Total Offspring	P value at 0.25	P value at 0.125
A261	3	34	0.02	0.37
A262	6	40	0.1	0.38
A262.1	1	9	0.3	0.69
A262.2	2	21	0.07	0.5
Total	12	104	0.0003	0.12

What can be seen here is that in the case of the del219F24 line A240, the numbers observed were more likely to be at a frequency of 0.25 than 0.125, while in the case of each of the 219F24 lines the observed numbers of wasted mice were more likely to occur at a frequency of 0.125 than 0.25.

It appears therefore that the presence of 219F24 in a homozygous wasted mouse is enough to rescue that mouse from phenotype. It should also be noted that no wasted mouse produced in these crosses was transgenic.

However the presence of del219F24 was not enough to rescue the wasted mice. Here there were several obviously wasted mice that were transgene positive and upon examination it was found that these animals exhibited the spleen and thymus atrophy apparent in non-transgenic wasted mice. It therefore seemed that the loss of functional eEF1A2 was solely responsible for the whole of the wasted phenotype.

It was important in these animals, however to ensure that this transgene was actually being expressed, as there was no difference between transgenic positive and negative mice. To do this a Western blot was carried out which compared the level of Clwd expression in transgenic and non-transgenic animals. As the transgene contained an intact copy of Clwd, if it was expressed, then there should at least be a 1.5-fold increase in Clwd expression in the transgenic animals given a copy-number of one. Using kidney protein extracts from both transgenic and non-transgenic animals, a Western blot was carried out using the anti-Clwd antibody and anti-α-Tubulin as a loading control. Kidney extracts were used as this tissue has never been implicated in any of the disease processes associated with wasted mice. The results can be seen in figure 6.4 in addition to graphs comparing the ratio of Clwd to α-Tubulin expression in each of the tissues examined. Using an unpaired t-test to examine how different these two sets of data are, it was found that the two-tailed P value equalled 0.0175 which, by conventional criteria, was a difference considered to be statistically significant. Looking at the average values for the mice examined it appears that Clwd expression in transgenic mice is approximately 1.5 times that of the non-transgenic mice and so suggests a copy-number of 1.

This implies that the transgene, although expressing Clwd, cannot rescue any aspect of the wasted phenotype and so that Clwd is not responsible for the pathology of wasted mice.

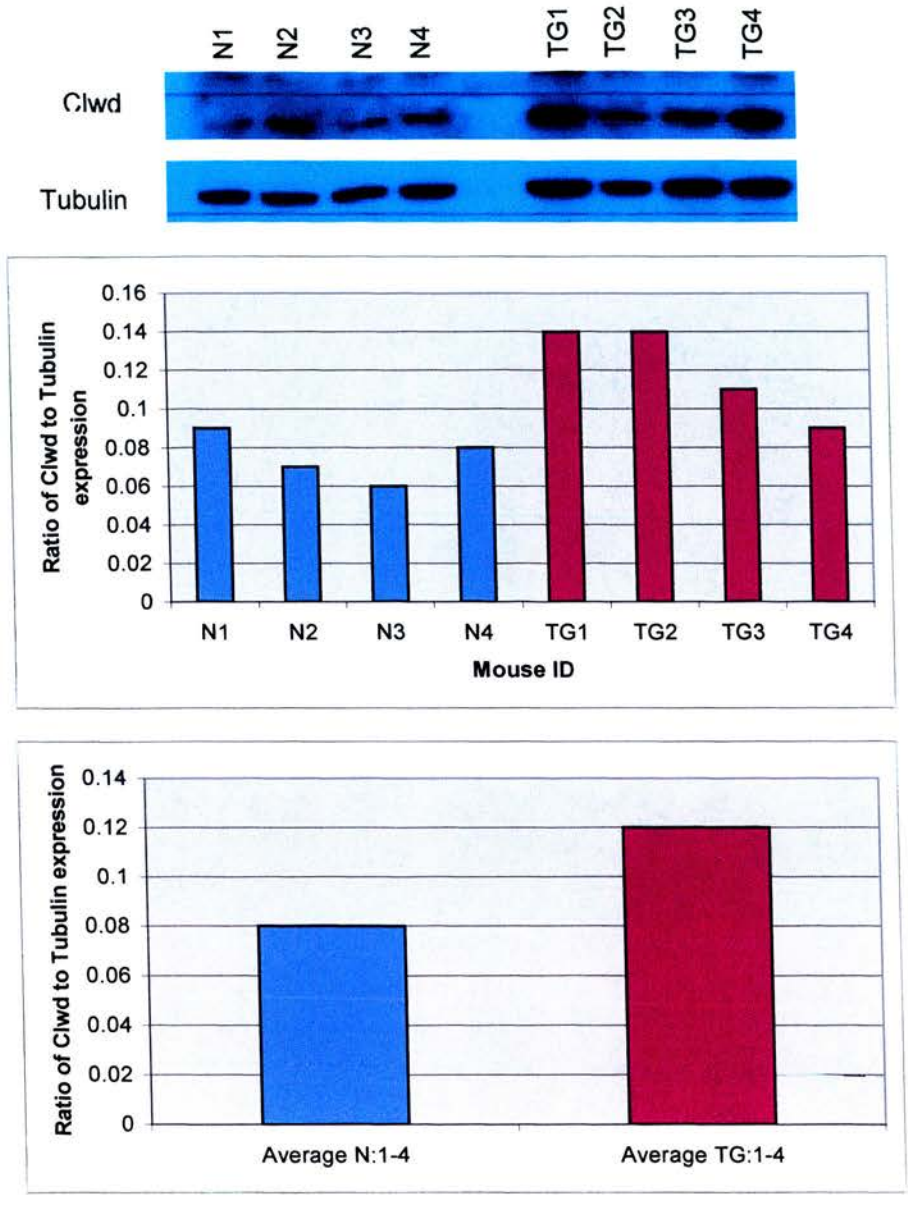


Figure 6.4 Comparison of Clwd expression in kidney protein extracts from wild type (N1-4) and transgenic (TG1-4) mice from the A240 line of del219F24.

6.2.2 Comparison of Clwd expression in wild type and wasted mice

Having an anti-Clwd antibody was extremely useful to compare the expression of Clwd in wild type and wasted mice. Western blots using this antibody showed that, for all the tissues examined, there did not appear to be any difference in expression of Clwd between *wst/wst* and wild type mice. The only difference at all appeared in the spleen where the wild type spleen exhibited a doublet whereas only a singlet was apparent in the wasted lane. Although there was evidence of alternative splice forms of Clwd, these are predicted to produce either full-length protein or one so truncated (20 amino acids long) that it is unlikely to be produced *in vivo*. I have speculated previously (see chapter 4) that this doublet is present due to alternatively phosphorylated forms of Clwd. If this is true it is unlikely that this is a major difference then between wild type and wasted spleen as phosphorylation is often a transient change to a protein and so is more likely to be relevant to the state of the tissue when it was acquired than any real major change in the expression of Clwd.

Figure 6.5 shows the Western blot comparison of wild type and wasted tissues.

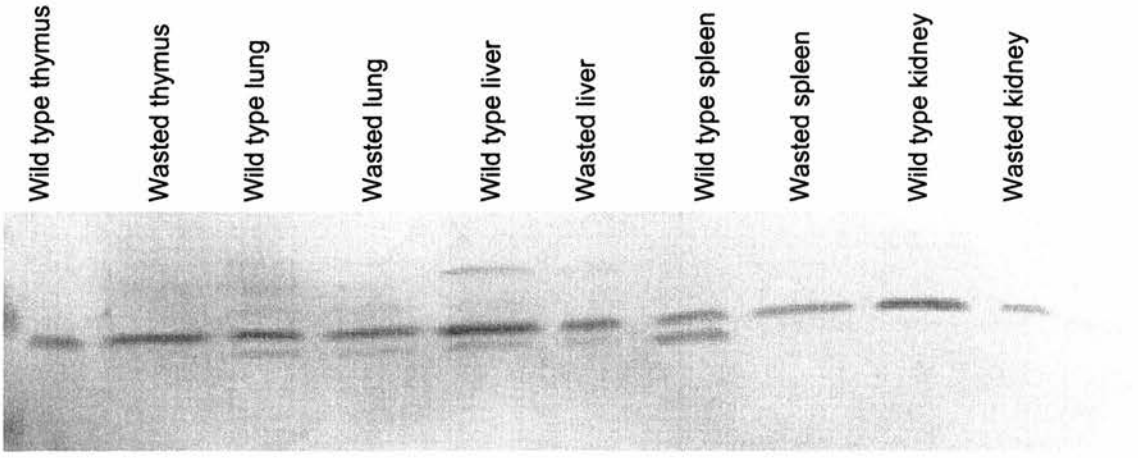


Figure 6.5 Clwd expression in wild type and wasted tissues. Using the anti-Clwd antibody on a Western blot there appears to be no difference in the expression of Clwd between wild type and wasted tissues. In the spleen a doublet is apparent in the wild type sample which is not seen in the wasted sample. This could be due to differences in the phosphorylation status of Clwd between the two tissues

6.3 Is eEF1A2 a good candidate? Expression in dendritic cells.

It now seemed that eEF1A2 was a better candidate for the immune system pathology of wasted mice than Clwd. As discussed in the introduction to this thesis, there was some evidence in the literature that eEF1A2 might be expressed in the dendritic cells of the immune system – even though this expression was not detectable in a whole immune system organ such as the spleen or thymus. In addition to this paper, an article describing Prof. Colin Watts' work at the University of Dundee drew our attention to the microarray data collected by his group examining the differences in expression profile between immature and mature dendritic cells[59]. Following personal communications with Prof Watts, it was discovered that according to these microarray data, *Eef1a1* expression decreased approximately 10-fold upon maturation of dendritic cells. As an apparent isoform-switch takes place in wild type muscle (*i.e.* upon the decrease in eEF1A1 post-natally eEF1A2 expression increases) it was possible that something similar was occurring upon the maturation of dendritic cells. To address this possibility the expression levels of *Eef1a1* and *Eef1a2* were monitored by RT-PCR on bone marrow-derived dendritic cell mRNA kindly donated by Prof. Watts. The mRNA came from dendritic cells before and after the LPS (LipoPolySaccharide) treatment that induces maturation. In addition to *Eef1a1* and *Eef1a2* the levels of three different maturation markers were also examined – *Fn1* (Fibronectin 1), *Ccl5* (Chemokine (C-C motif) ligand 5) and *Il1b* (Interleukin 1 beta). On maturation of dendritic cells, expression of *Fn1* decreases while that of *Ccl5* and *Il1b* increases. Figure 6.5 shows the RT-PCR experiment examining the expression of these genes. As can be seen from this result, the maturation markers behave as expected. Both *Eef1a1* and *Eef1a2* are expressed in dendritic cells with *Eef1a1* expression dramatically decreasing on maturation while *Eef1a2* expression increases slightly.

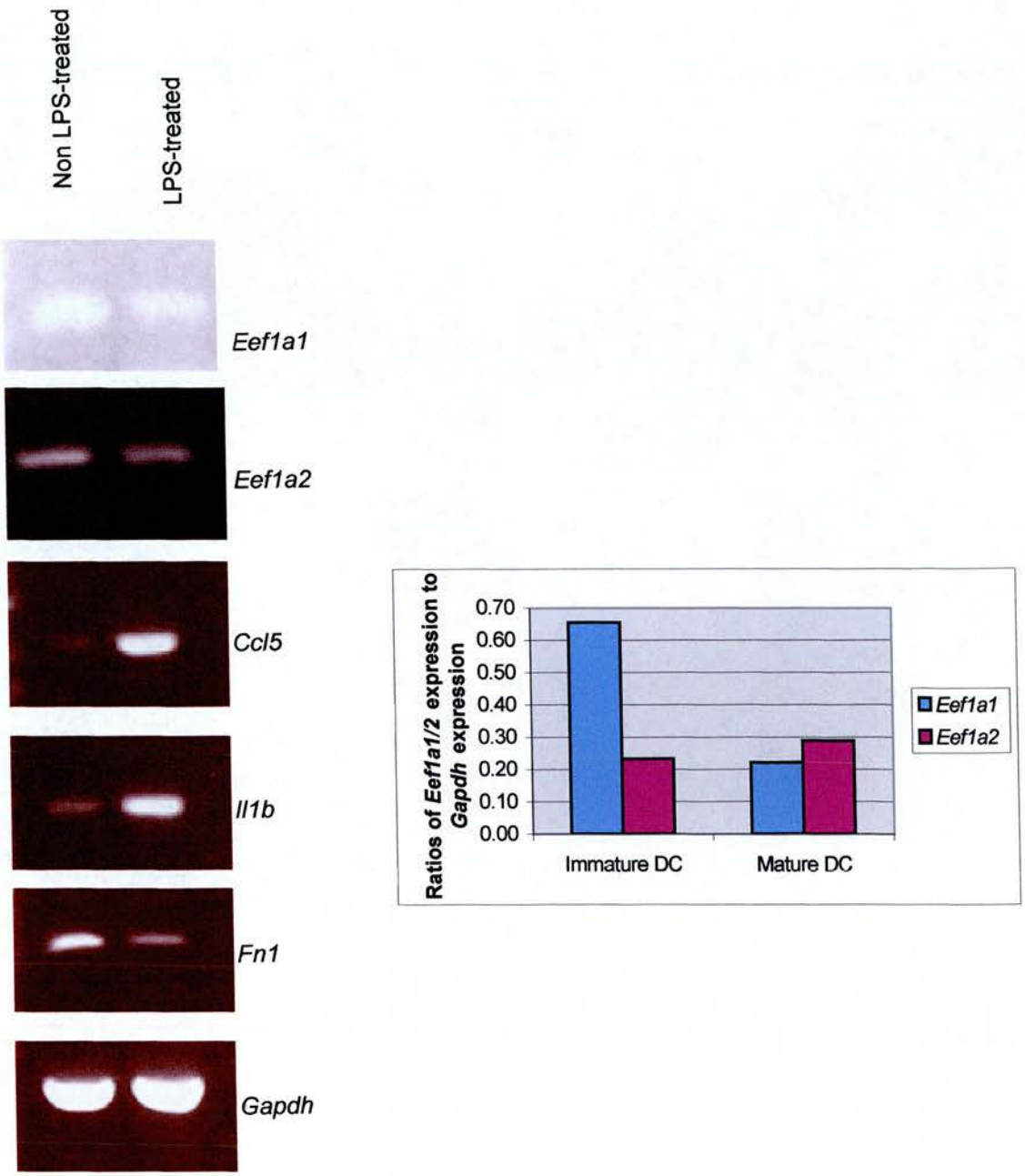


Figure 6.6 *Eef1a1*, *Eef1a2* and maturation marker expression by RT-PCR in dendritic cells with and without LPS treatment

6.4 Discussion.

Following the comparison of Clwd expression in wild type and wasted mice and in particular the transgenic experiment described here, it became apparent that the likelihood of the wasted deletion having a functional impact on Clwd was slim. The transgenic experiment also suggested it was only the wasted deletion's effect upon *Eef1a2* that resulted in the total pathology associated with the wasted mice. As this became apparent, a preliminary investigation into the expression of *Eef1a2* in the dendritic cells of the immune system was conducted and it was found that *Eef1a2* was indeed expressed in the dendritic cells. It was also discovered that on maturation of the dendritic cell there was a sharp decrease in *Eef1a1* expression. This could mean that in mature dendritic cells eEF1A2 is of increased importance. How a loss of eEF1A2 expression, as in wasted mice, would affect the dendritic cell population is unclear and how this affected population of cells could cause the immune system abnormalities observed in wasted mice is unknown. This work does serve to prove however that not only is *Eef1a2* expressed in a component of the immune system, but also it appears its expression is vital not only to that component but the whole system. Perhaps in future a conditional knockout of *Eef1a2* in the immune system would allow its role in this system to be assessed in more detail

CHAPTER 7

DISCUSSION

7. Discussion

7.1 Aims and Methodology

The aim of this project was to test if a novel gene discovered in the vicinity of the wasted deletion was involved in the immune system abnormalities observed in the mouse mutant wasted. Having determined at the start of this project that this novel gene was a good candidate to study, I decided first to characterise this gene and its protein product before addressing its potential role in the pathology of the wasted immune system. I planned to do this by conducting an as detailed as possible examination of the novel gene Clwd *in silico* in order to inform my future *in vitro* work. Subsequently I hoped to use this information to confirm that this gene, predicted by various gene prediction programmes did produce an mRNA transcript and protein product. Finally I planned to use the information collated thus far to elucidate the function of the protein product of this gene.

Having completed this characterisation of the gene, I hoped to address whether or not it played a role in the wasted phenotype. This would be done by examining the results of a transgenic experiment in addition to using my new knowledge of Clwd and its features to assess the part it played, if any in this pathology.

7.2 The novel gene Clwd *in silico* and *in vitro*

Clwd was initially uncovered by my group through the careful cloning and sequencing associated with mapping a genetic lesion resulting from a classical mutation. In this case the lesion was the wasted deletion which removes 15.8kb of distal mouse chromosome 2. It was seen originally as a portion of the sequence with an apparent CpG island and homology to a large number of ESTs.

Two advances during the course of the project made the study of Clwd as a novel gene easier. Firstly the mouse genome was in the process of, and then finally sequenced[50] and with this came an emphasis on gene prediction programmes to annotate new genes by learning from the structure of known genes. Secondly the sequence of the mouse transcriptome became available with the advent of full-length mouse cDNA clones[60]. These two advances made my study of Clwd as a predicted gene far easier and allowed me to use the wealth of new software online to predict its potential structure and function. From these I had confirmed Clwd genome and mRNA sequence which I used to elucidate the intron-exon structure of the gene.

From the alignment of these sequences I could determine that Clwd appeared to be alternatively spliced. This alternative splicing took the form of skipped exons as in the majority of alternative splicing events, but also a combination of the use of alternative 5'and 3' donor sites in the first and second exons, and the retention of the first and second introns, the latter being relatively rare in known alternatively spliced genes. There were 5 splice forms in all.

Most of the alternatively splice locations occurred in the 5'UTR. This is not unusual as can be seen in figure 7.1. It is thought that alternative splicing in the 5'UTR can lead to alterations in the stability and/or localisation of the mature mRNA without changing the sequence of the protein. However the splice location of major interest, that could really affect the production of a functional Clwd protein, was that which occurred within the coding sequence *i.e.* the retention of the second intron.

Due to the retention of intron 2 in splice form II, IV and V, a premature stop codon would be predicted to occur in the transcript from this intronic sequence. If true, this would have several implications. In most cases if a termination codon appears more than 50 nucleotides upstream of the final exon it is a PTC (Premature Termination Codon) and the mRNA that harbours it will be degraded by nonsense-mediated decay (NMD)[61]. This is a form of RNA quality control which prevents the translation of potentially dominant negative, C-terminal truncated proteins. This type of regulated splicing to produce unproductive isoforms is also used as a means of down regulating protein expression, as these mRNA isoforms remain untranslated. In *C. elegans* for example, the expression of the ribosomal proteins L3, L7a, L10a and L12 are regulated post-transcriptionally via the coupling of alternative splicing and NMD[62].

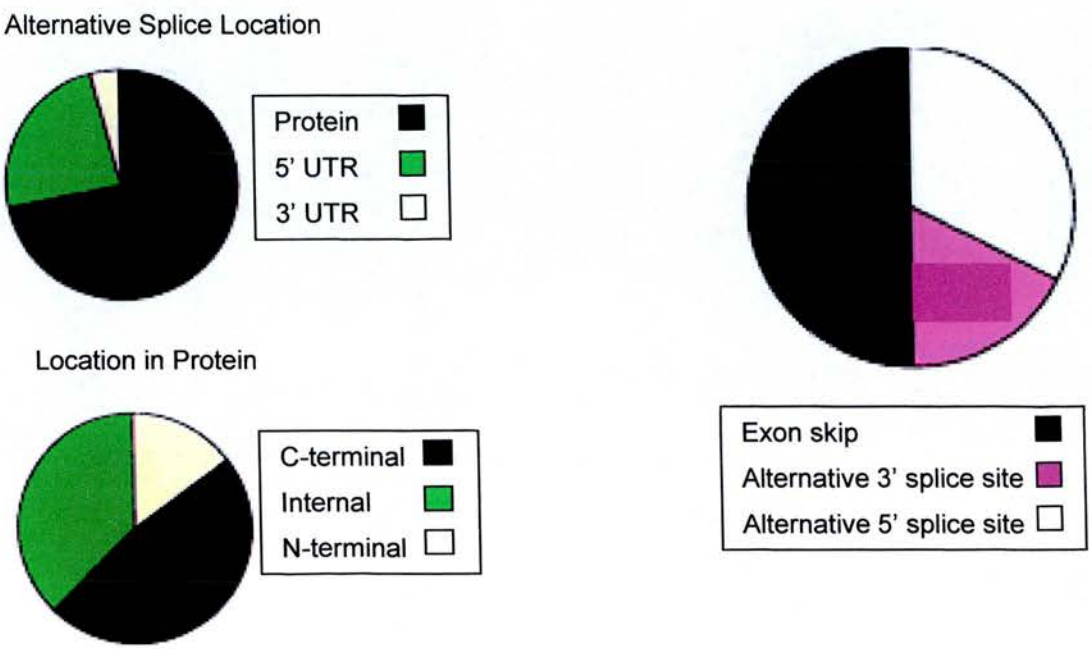


Figure 7.1 Pie charts depicting prevalence of various forms of alternative splicing in a random sample of alternatively spliced genes in human (from ref[63])

Looking at the expression of the splice forms mouse tissues by RT-PCR it appeared that the splice forms representing the retention of intron 2 were the least represented in each of the tissue examined. The other type of splice form, *i.e.* where the splice occurred in the 5'UTR, were more highly expressed. This would seem to suggest again that while changes to the 5'UTR can affect the transcript, it is the presence of a PTC that would have a more devastating effect.

It is worth noting however that the proportion of Clwd mRNA retaining this second intron is dramatically increased in the spleen sample examined. Why this occurs is unclear. Is it that in the spleen there is increased production of the forms retaining intron 2, or rather that these forms survive NMD by some mechanism? There has been some recent evidence that NMD can be tissue-specific in nature. Specifically, a Schmid metaphyseal chondrodysplasia patient with a mutation in the gene for collagen X resulted in a transcript with a PTC. This led to complete NMD of the transcript and thus collagen X haploinsufficiency in the cartilage of this patient. However the mutant mRNA has been found in non-cartilage cells – lymphoblasts and bone cells – and is not subject to NMD[64]. This would suggest that RNA surveillance and NMD can be regulated in some way and so the presence of a PTC in a transcript does not necessarily mean that transcript is degraded by NMD.

From my results there appeared to be a variety of splice variants present in each tissue. It is worth noting however that these images are just snapshots of the tissue at the time of retrieval. Clwd may exert its effect on a more cellular level such that in some cell types form I is more prevalent, where in others form V is the dominant form for example. Another possibility, if all cells require Clwd, is that different splice forms are required for different stages of the cell cycle.

The presence of Clwd transcripts with different 5'UTR structures could also influence its eventual expression levels by an alternate mechanism *i.e.* through the upstream Open Reading Frame or uORF in the 5'UTR. In contrast to alternative splicing, uORFs can control expression of the main ORF by repressing its translation.

By regulated alternative splicing and translation repression, Clwd's expression at both the transcription and translation level may be tightly controlled. Given that I have seen no difference in Clwd expression on a tissue level, this control could be at a more cellular level.

7.3 Clwd as a protein *in silico* and *in vitro*

What made Clwd interesting, in addition to a potential role in the wasted phenotype, was its novelty. The 114 amino acid sequence that is predicted to form the Clwd protein is unlike any known protein family. It contains no motifs that would allow a function to be assigned to it and it has not been described previously in the literature. It appears that despite our extensive knowledge of the mouse genome, there are functional genes about which we know nothing. More precisely of the 29,201 genes predicted to be encoded in the mouse genome, 9,785 are of unknown function.

Using motif recognition software the only functional motif Clwd appears to have, other than a host of post-translational modifications, is a form of GTP/ATP binding site called a P loop. On the surface therefore it would seem, given that Clwd is a small protein, perhaps it might be a good candidate for a new member of the small G protein family as they are monomeric proteins with molecular masses of 20-40 kD. However looking at the GTP-binding proteins more carefully it becomes clear that the "P-loop" is one domain of the four required to bind and hydrolyse GTP. The function of the P-loop fold in the protein is simply to position the triphosphate moiety of a bound nucleotide[65].

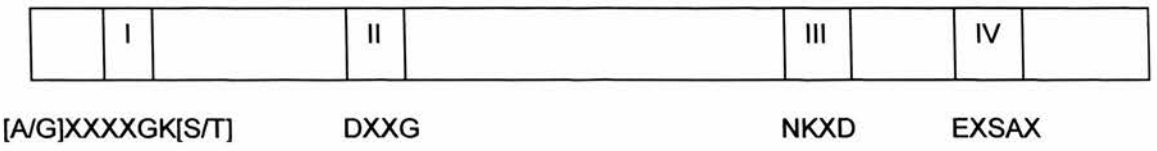


Figure 7.2 Protein structure of small GTP-binding proteins. Motif I is the P loop – the only motif possessed by Clwd.

Clwd is therefore unlikely to be a small G protein *i.e.* belonging to the Ras superfamily as it does not possess any of these other domains required to function as a GTPase. However the fact that it does contain a sequence that could potentially fold into a conformation that can interact with an ATP or GTP molecule could still be important for an unrecognised function.

When the actual Clwd protein was examined by Western blot it was found that its expression was as ubiquitous as the mRNA. The two main points worth noting were that Clwd's apparent size on the blot was approximately 28kD and that in some tissues there appeared to be a doublet. I believe both these features could be explained if Clwd has post-translational modifications. Clwd does have the motifs for a range of modifications such as O-GlcNAc and phosphate groups.

7.4 Location and function of Clwd

7.4.1 Sub-cellular localisation of Clwd

Upon trying to elucidate the molecular function of Clwd, the first step to take was to locate the protein within the cell. This was done by the co-localisation of various organelle markers with Clwd and I found that the best co-localisation here was with a protein Paxillin.

Paxillin is found at the site of focal adhesions connecting the extracellular matrix to the actin cytoskeleton. Adhesion of cells to the extracellular matrix is fundamental for normal cellular homeostasis. Signals are transduced through the integrin family of matrix receptors. These, in turn, activate intracellular signalling cascades that regulate the actin cytoskeleton network and the gene expression changes required for embryogenesis, normal cell proliferation, survival and wound repair. The specialised sites of cell contact formed between the plasma membrane and the extracellular matrix, called focal adhesions comprise multiple structural and signalling proteins. These proteins are associated with the cytoplasmic domains of integrins subunits necessary for coordinating these signalling events[66]. Paxillin is a focal adhesion associated adaptor protein that recruits signalling molecules into a complex[67].

Although co-localisation experiments with Golgi markers did not give clear results Clwd did have similar staining to Vti1a/b in NIH/3T3 cells, Golgi proteins involved in vesicle trafficking. Given these results it is worth noting that there is a family of small G proteins that interact with Paxillin and yet are also found in the perinuclear region of the Golgi. These are termed the ARF-GAPs. In fact focal adhesion proteins like Paxillin and Vinculin have also been found to co-exist with various ARF GAPs in both the perinuclear Golgi region and at the plasma membrane[66]. If Clwd functions in one of the signalling pathways tied to Paxillin therefore, it would not be unusual to see both perinuclear and more widespread cytoplasmic staining.

7.4.2 Silencing of the Clwd gene in mammalian cells

As an initial experiment to assess Clwd's function in cells I decided to use RNAi to silence the gene in a mouse fibroblast (NIH/3T3) cell line. This approach gave some interesting results that suggested Clwd might be important in cell and/or nuclear architecture or in normal cell growth and survival. However I found results using this method to be inconsistent. Using chemically synthesised siRNA molecules and transfecting them into cells to obtain a short-term silencing effect was the first method described to reproduce RNAi in mammalian cell culture. However since this was first discovered[55], other methods of gene silencing by RNA interference have been developed which may prove of more use in the future. In particular siRNA expression vectors have been developed where a plasmid can be stably transfected into cells and then be induced to produce hairpin siRNAs. This means the transfection efficiency is no longer a problem as cells transfected can be selected for, before induction of gene silencing. In addition, long term silencing can be achieved rather than the 4-5 day silencing seen in using synthesised siRNAs.

As RNAi in cell lines is a relatively new methodology, new techniques and resources are being developed all the time to make it more efficient. Therefore the inconsistency seen here by using chemically synthesised siRNAs would not dissuade me from using RNAi in the future to silence Clwd in cells but I would choose another method

7.5 Wasted immune system

The original reason for studying Clwd was in order to assess whether or not it was a good candidate for playing a part in the phenotype of the mouse mutant wasted. As discussed in the introduction to this thesis I believed that Clwd was the best candidate to be the gene responsible for the immune system phenotype for a number of reasons. Therefore a lot of effort was spent in characterising Clwd as it was a novel gene, not described in the literature, and also unlike any other gene currently known.

However upon a combined use of transgenic animals to rescue the wasted phenotype and comparing Clwd expression in wild type and wasted mice, it became clear that not only was Clwd not responsible for the wasted phenotype, but the loss of *Eef1a2* expression accounted for all pathologies associated with wasted mice.

As discussed in the introduction, one paper had described *Eef1a2* expression in dendritic cells[37]. I therefore chose to confirm this expression in bone marrow derived dendritic cells by RT-PCR and found not only is *Eef1a2* expressed in dendritic cells, but upon maturation of the dendritic cell there is a dramatic decrease in *Eef1a1* expression while *Eef1a2* expression increases, but only slightly. This implies that in a wasted mouse the dendritic cells would have little or no eEF1A expression at all. How this would affect the dendritic cell is uncertain.

As these dendritic cells express both eEF1A1 and eEF1A2 the fact that the loss of eEF1A2 can have such a devastating effect implies that their individual functions in dendritic cells are not equivalent. Alternatively it could be the reduction in the amount of total eEF1A in wasted dendritic cells that causes the immune system defects. However it is worth noting that while in the mouse eEF1A1 and eEF1A2 are almost mutually exclusively expressed by cell type (except for some evidence of eEF1A1 expression in the nuclei of spinal cord motor neurons), in most cultured cell lines examined both are expressed. Therefore the presence of eEF1A1 in the dendritic cells examined could be an artefact of cell culture and if one looked at

dendritic cells *in vivo* only eEF1A2 would be expressed. If this were the case, then in wasted dendritic cells there would be no eEF1A whatsoever and this could be causing the effects seen in the wasted immune system.

How the loss of a translation elongation factor and/or a cytoskeleton interacting protein in one component of the immune system can have such an affect on other cells of the immune system – particularly in their ability to proceed through the cell cycle with all the correct checkpoints is unclear. However what is clear is that the loss of eEF1A2 does cause the immune system pathology. How this happens now needs to be addressed.

7.6 Future Work

7.6.1 Clwd

Clwd is a novel, and at the moment, unique gene in both the human and mouse (and other) genomes. As such, the amount of work to be done on it and its protein product is almost endless at this stage. However some questions I would be very interested in addressing are as follows.

It appears from my experiments that all of the alternatively spliced transcripts of Clwd are expressed in every tissue examined. As I believe that the splice forms and the uORF have some sort of regulatory function, albeit not a tissue-specific regulation, I would like to examine the expression of these splice forms and the protein at discrete points during the cell cycle. In mammalian cell culture there are various drugs and methods of synchronising cells and stopping them in various stages of the cell cycle. Examining Clwd expression in these stages may shed some light on the reason for the different splice forms seen. I also believe it important to confirm that the uORF is functional as a translation repressor, and if so, would the different splice forms exert different levels of repression.

Given the unclear results following the Northern blot of the matched normal and tumour tissue array I would like to carry out splice form-specific hybridisation experiments using these sample to see if it is the increase or decrease in a particular isoform that occurs during tumourigenesis

Turning to the protein product of Clwd, I would like to confirm that the sequence predicted is indeed the true amino acid sequence of Clwd. I think the protein should be purified and sequenced in addition to raising multiple antibodies against Clwd using bacterially expressed protein and different peptide sequences for the creation of anti-peptide antibodies. In addition any post-translational modifications need to be mapped as these could also elucidate its function.

In order to assess Clwd's function there are a variety of approaches to be taken. As any interactions with known proteins are impossible to predict, given Clwd's lack of homology to any known protein, an insight into its function could be gained from using a yeast two hybrid screen to at least place Clwd within some system in the cell.

As Clwd is so widely and early expressed, I would anticipate that a straight-forward knockout of Clwd may be embryonic-lethal and so I would instead create a conditional knockout so that the loss of functional Clwd could be controlled temporally and spatially to gain the maximum amount of information possible. However I also anticipate, again given its expression pattern, that Clwd serves its function at a cellular level. Therefore a long-term gene silencing experiment in mammalian cell lines using RNAi expression vectors could reveal just as much, if not more, information than a mouse knockout.

These are just the first round of experiments that I would carry out on Clwd next, with their results informing and directing a more in depth study of Clwd and its function.

7.6.2 The immune system pathology of wasted mice

The discovery that the loss of *Eef1a2* expression is responsible for the immune system phenotype is just the first step in elucidating this pathology. In order to address eEF1A2's role here I would first check to see if it is expressed in any other immune system cell types. Although no expression is seen in spleen and thymus as a whole, I believe it would be important to ascertain that the effect of the loss of eEF1A2 expression is mediated through the dendritic cells alone. Additionally it would be important to look at *Eef1a1* expression in native dendritic cells rather than those derived from bone marrow and matured in culture to ensure its expression is not an artefact of cell culture.

Because the neuromuscular deficiencies in *wst/wst* mice are so severe and kill the mice by 28 days of age, if there was some way of separating out the two phenotypes it may be possible to determine the progression of the immune system pathology. I believe this could be done by the creation of bone marrow chimaeras where the bone marrow of a wasted mouse could be transferred into a wild type mouse whose active bone marrow had been removed by radiation. As the dendritic cells arise in the bone marrow, the wild type mouse with wasted bone marrow would be expected to have the immune deficiencies of the wasted mice but with a healthy neuromuscular system allowing them to live longer so the *wst/wst* immune system could be studied in detail. Alternatively, a conditional *Eef1a2* knockout could be created where *Eef1a2* is only removed from the immune system. This would again be predicted to allow the mice to live longer and so their immune system could be studied in isolation.

7.7 Final Comments

This thesis, in addressing the question of which gene is responsible for the immune system defects in the mouse mutant wasted has generated questions as well as answers. It has answered the original question - we now know the loss of eEF1A2 is responsible for these defects. However it has introduced the novel gene, Clwd and while headway has been made in characterising this gene and its protein product many questions remain to be answered: What is the function of Clwd? What proteins does it interact with? How is it regulated?
In addition, although we know the loss of eEF1A2 has a devastating effect on the spleen and thymus of wasted mice, we do not know how this effect is mediated.

I hope that these questions will be answered in the months and years to come and that the answers will enlighten some of the fundamental mechanisms of cellular function and immunological health.

BIBLIOGRAPHY

1. Shultz, L.D., et al., 'Wasted', a new mutant of the mouse with abnormalities characteristic to ataxia telangiectasia. Nature, 1982. 297(5865): p. 402-4.

2. Lutsep, H.L. and M. Rodriguez, Ultrastructural, Morphometric, and Immunocytochemical Study of Anterior Horn Cells in Mice with "Wasted" Mutation. J Neuropathol Exp Neurol, 1989. 48(5): p. 519-33.

3. Gatti, R.A., et al., Localization of an ataxia-telangiectasia gene to chromosome 11q22-23. Nature, 1988. 336(6199): p. 577-80.

4. Siracusa, L.D., et al., Mouse Chromosome 2. Mamm Genome, 1997. 7(Supp): p. S28-44.

5. Savitsky, K., et al., A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science, 1995. 268(5218): p. 1749-53.

6. Chambers, D.M., J. Peters, and C.M. Abbott, The lethal mutation of the mouse wasted (wst) is a deletion that abolishes expression of a tissue-specific isoform of translation elongation factor 1alpha, encoded by the Eef1a2 gene. Proc Natl Acad Sci U S A, 1998. 95(8): p. 4463-8.

7. Nordeen, S.K., et al., Evaluations of wasted mouse fibroblasts and SV-40 transformed human fibroblasts as models of ataxia telangiectasia in vitro. Mutat Res, 1984. 140(4): p. 219-22.

8. Tezuka, H., et al., Evaluation of the mouse mutant "wasted" as an animal model for ataxia telangiectasia. I. Age-dependent and tissue-specific effects. Mutat Res, 1986. 161(1): p. 83-90.

9. Inoue, T., et al., Effect of DNA-damaging agents on isolated spleen cells and lung fibroblasts from the mouse mutant "wasted," a putative animal model for ataxia-telangiectasia. Cancer Res, 1986. 46(8): p. 3979-82.

10. Hecht, S.M., Bleomycin: new perspectives on the mechanism of action. J Nat Prod, 2000. 63(1): p. 158-68.

11. Fronza, G., et al., The 4-nitroquinoline 1-oxide mutational spectrum in single stranded DNA is characterized by guanine to pyrimidine transversions. Nucleic Acids Res, 1992. 20(6): p. 1283-7.

12. van Buul, P.P., et al., A search for radiosensitive mouse mutants by use of the micronucleus technique. Mutat Res, 1987. 191(3-4): p. 163-9.

13. van Buul, P.P., et al., Cytogenetic characterization of radiosensitive mouse mutants. Mutat Res, 1991. 251(2): p. 171-9.

14. **Libertin, C.R., et al., Dysregulation of temperature and liver cytokine gene expression in immunodeficient wasted mice. Cell Immunol, 1996. 169(1): p. 62-6.**

15. **Libertin, C.R., et al., Subnormal albumin gene expression is associated with weight loss in immunodeficient/DNA-repair-impaired wasted mice. J Am Coll Nutr, 1994. 13(2): p. 149-53.**

16. **Woloschak, G.E., et al., Regulation of thymus PCNA expression is altered in radiation-sensitive wasted mice. Carcinogenesis, 1996. 17(11): p. 2357-65.**

17. **Potter, M., A. Bernstein, and J.M. Lee, The *wst* Gene Regulates Multiple Forms of Thymocyte Apoptosis. Cellular Immunology, 1998. 188(2): p. 111-117.**

18. **Ann, D.K., et al., Isolation and characterization of the rat chromosomal gene for a polypeptide (pS1) antigenically related to statin. J Biol Chem, 1991. 266(16): p. 10429-10437.**

19. **Abbott, C.M. and C.G. Proud, Translation factors: in sickness and in health. Trends Biochem Sci., 2004. 29(1): p. 25-31.**

20. **Lee, S., et al., Tissue-specific expression in mammalian brain, heart, and muscle of S1, a member of the elongation factor-1 alpha gene family. J Biol Chem, 1992. 267(33): p. 24064-24068.**

21. **Lund, A., et al., Assignment of human elongation factor 1alpha genes: EEF1A maps to chromosome 6q14 and EEF1A2 to 20q13.3. Genomics, 1996. 36(2): p. 359-361.**

22. **Kahns, S., et al., The elongation factor 1 A-2 isoform from rabbit: cloning of the cDNA and characterization of the protein. Nucleic Acids Res, 1998. 26(8): p. 1884-90.**

23. **Khalyfa, A., et al., Characterization of elongation factor-1A (eEF1A-1) and eEF1A-2/S1 protein expression in normal and wasted mice. J Biol Chem, 2001. 276(25): p. 22915-22922.**

24. **Feig, S. and P. Lipton, Pairing the cholinergic agonist carbachol with patterned Schaffer collateral stimulation initiates protein synthesis in hippocampal CA1 pyramidal cell dendrites via a muscarinic, NMDA-dependent mechanism. J Neurosci., 1993. 13(3): p. 1010-21.**

25. **Ruest, L.B., R. Marcotte, and E. Wang, Peptide elongation factor eEF1A-2/S1 expression in cultured differentiated myotubes and its protective effect against caspase-3-mediated apoptosis. J Biol Chem, 2002. 277(7): p. 5418-25.**

26. **Dharmawardhane, S., et al., Compartmentalization and actin binding properties of ABP-50: the elongation factor-1 alpha of Dictyostelium. Cell Motil Cytoskeleton, 1991. 20(4): p. 279-88.**

27. **Munshi, R., et al., Overexpression of translation elongation factor 1A affects the organization and function of the actin cytoskeleton in yeast. Genetics, 2001. 157(4): p. 1425-36.**

28. **Moore, R.C., N.A. Durso, and R.J. Cyr, Elongation Factor-1a Stabilizes Microtubules in a Calcium/Calmodulin-Dependent Manner. Cell Motil Cytoskeleton, 1998. 41(2): p. 168-180.**

29. **Shiina, N., et al., Microtubule severing by elongation factor 1 alpha. Science, 1994. 266(5183): p. 282-5.**

30. **Ohta, K., et al., The mitotic apparatus-associated 51-kDa protein from sea urchin eggs is a GTP-binding protein and is immunologically related to yeast polypeptide elongation factor 1 alpha. J Biol Chem, 1990. 265(6): p. 3240-7.**

31. **Owen, C.H., D.J. DeRosier, and J. Condeelis, Actin crosslinking protein EF-1a of Dictyostelium discoideum has a unique bonding rule that allows square-packed bundles. J Struct Biol., 1992. 109(3): p. 248-54.**

32. **Bassell, G.J., et al., Single mRNAs visualized by ultrastructural in situ hybridization are principally localized at actin filament intersections in fibroblasts. J Cell Biol, 1994. 126: p. 863-876.**

33. **Derventzi, A., S.I. Rattan, and B.F. Clark, Phorbol ester PMA stimulates protein synthesis and increases the levels of active elongation factors EF-1 alpha and EF-2 in ageing human fibroblasts. Mech Ageing Dev., 1993. 69(3.): p. 193-205.**

34. **Edmonds, B.T., et al., Elongation factor-1 alpha is an overexpressed actin binding protein in metastatic rat mammary adenocarcinoma. J Cell Sci, 1996. 109(11): p. 2705-14.**

35. **Duttaroy, A., et al., Apoptosis rate can be accelerated or decelerated by overexpression or reduction of the level of elongation factor-1 alpha. Exp Cell Res, 1998. 238(1): p. 168-76.**

36. **Chen, E., et al., Rapid up-regulation of peptide elongation factor EF-1alpha protein levels is an immediate early event during oxidative stress-induced apoptosis. Exp Cell Res, 2000. 259(1): p. 140-8.**

37. **Sandberg, J.K., et al., T cell tolerance based on avidity thresholds rather than complete deletion allows maintenance of maximal repertoire diversity. J Immunol, 2000. 165(1): p. 25-33.**

38. Flores-Romo, L., In vivo maturation and migration of dendritic cells. Immunology, 2001. 102(3): p. 255-62.

39. Salomon, B., et al., Conditional ablation of dendritic cells in transgenic mice. J Immunol, 1994. 152(2): p. 537-48.

40. Binks, M., et al., Intrinsic dendritic cell abnormalities in Wiskott-Aldrich syndrome. Eur J Immunol, 1998. 28(10): p. 3259-67.

41. Symons, M., et al., Wiskott-Aldrich syndrome protein, a novel effector for the GTPase CDC42Hs, is implicated in actin polymerization. Cell, 1996. 84(5): p. 723-34.

42. Snapper, S.B., et al., Wiskott-Aldrich Syndrome Protein-Deficient Mice Reveal a Role for WASP in T but Not B Cell Activation. Immunity, 1998. 9: p. 81-91.

43. Kleinjan, D.J. and V. van Heyningen, Position effect in human genetic disease. Hum Mol Genet, 1998. 7(10): p. 1611-8.

44. Tinel, N., et al., The KCNQ2 potassium channel: splice variants, functional and developmental expression. Brain localization and comparison with KCNQ3. FEBS Lett., 1998. 438(3): p. 171-6.

45. Watanabe, H.N., E; Kosakai, A; Nakamura, M; Yokoyama, M; Tanaka, K; Sasai, H, Disruption of the epilepsy KCNQ2 gene results in neural hyperexcitability. J Neurochem, 2000. 75(1): p. 28-33.

46. Vasioukhin, V., et al., A novel intracellular epithelial cell tyrosine kinase is expressed in the skin and gastrointestinal tract. Oncogene. 10(2): p. 349-57.

47. Siyanova, E.Y., et al., Tyrosine kinase gene expression in the mouse small intestine. Oncogene, 1994. 9(7): p. 2053-7.

48. Kohmura, N., et al., A novel nonreceptor tyrosine kinase, Srm: cloning and targeted disruption. Mol Cell Biol, 1994. 14(10): p. 6915-25.

49. Elbashir, S.M., et al., Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods, 2002. 26(2): p. 199-213.

50. Consortium., M.G.S., Initial sequencing and comparative analysis of the mouse genome. Nature, 2002. 420(6915): p. 520-62.

51. Castillo-Davis, C.I., et al., Selection for short introns in highly expressed genes. Nat Genet, 2002. 31(4): p. 415-8.

52. Kondrashov, F.A. and E.V. Koonin, Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. Trends Genet, 2003. 19(3): p. 115-9.

53. **Geballe, A.P. and D.R. Morris, Initiation codons within 5'-leaders of mRNAs as regulators of translation. Trends Biochem Sci, 1994. 19(4): p. 159-64.**

54. **Paddison, P.J., A.A. Caudy, and G.J. Hannon, Stable suppression of gene expression by RNAi in mammalian cells. Proc Natl Acad Sci U S A, 2002. 99(3): p. 1443-8.**

55. **Elbashir, S.M., et al., Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature, 2001. 411(6836): p. 494-8.**

56. **Jackson, A.L., et al., Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol., 2003. 21(6): p. 635-7.**

57. **Chi, J.T., et al., Genomewide view of gene silencing by small interfering RNAs. Proc Natl Acad Sci U S A, 2003. 100(11): p. 6343-6.**

58. **Semizarov, D., et al., Specificity of short interfering RNA determined through gene expression signatures. Proc Natl Acad Sci U S A, 2003. 100(11): p. 6347-52.**

59. **GN, *On an Epic Scale: Studying proteins in Scotland*, in *Wellcome News*. 2001. p. 18-21.**

60. **Team*, T.F.C.a.t.R.G.E.R.G.P.I.I., Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature, 2002. 420: p. 563-573.**

61. **Hillman, R.T., R.E. Green, and S.E. Brenner, An unappreciated role for RNA surveillance. Genome Biol, 2004. 5(2): p. R8:1-15.**

62. **Mitrovich, Q.M. and P. Anderson, Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in C. elegans. Genes Dev, 2000. 14(17): p. 2173-84.**

63. **Modrek, B., et al., Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res, 2001. 29(13): p. 2850-2859.**

64. **Bateman, J.F., et al., Tissue-specific RNA surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in Schmid metaphyseal chondrodysplasia cartilage. Hum Mol Genet, 2003. 12(3): p. 217-225.**

65. **Takai, Y., T. Sasaki, and T. Matozaki, Small GTP-binding proteins. Physiol Rev, 2001. 81(1): p. 153-208.**

66. **Turner, C.E., K.A. West, and M.C. Brown, Paxillin-ARF GAP signlling and the cytoskeleton. Curr Opin Cell Biol, 2001. 13: p. 593-599.**

67. **Schaller, M.D., Paxillin: a focal adhesion-associated adaptor protein. Oncogene, 2001. 20: p. 6459-6472.**